умовне математичне сподівання $a_{y/x}$, що є деякою функцією від x: $a_{y/x} = g(x)$, називають ***регресією Y на X***.

Аналітичні вирази $a_{x/y} = h(y)$ і $a_{y/x} = g(x)$ називають ***рівняннями регресії***. Графіки функцій $a_{x/y} = h(y)$ і $a_{y/x} = g(x)$ називають ***лініями регресії***.

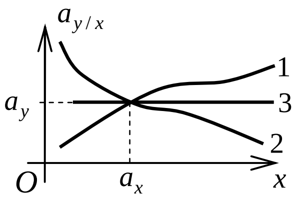


Рис. 33

Регресійний аналіз дозволяє виявити характер залежності між величинами X і Y. На рис. 33 лінія регресії 1 показує, що між X і Y існує додатна кореляційна залежність, при цьому $r_{xy} > 0$. Лінія регресії 2 показує, що між X і Y існує від'ємна кореляційна залежність, при цьому $r_{xy} < 0$. А лінія регресії 3 показує, що величини X і Y лінійно незалежні, при цьому $r_{xy} = 0$.

Якщо лінії регресії є прямими, то їх рівняння мають вигляд:

$$\boxed{a_{y/x} = Ax + B}; \quad \boxed{a_{x/y} = Cy + D},$$

де $\boxed{A = r_{xy}\,\sigma_y / \sigma_x}$; $\boxed{B = a_y - A a_x}$; $\boxed{C = r_{xy}\,\sigma_x / \sigma_y}$; $\boxed{D = a_x - C\,a_y}$.

Прямі регресії проходять через центр (a_x, a_y) сумісного розподілу величин X і Y.

1.5. Закон великих чисел. Граничні теореми

Нехай у ході статистичного експерименту визначаються значення деякої випадкової величини X. При однократному випробуванні не можна заздалегідь передбачити, яке з можливих значень прийме величина X. Проте при багатократному повторенні досліду середнє арифметичне величини X майже втрачає випадковий характер і наближається до деякого сталого значення.

Закон великих чисел – це сукупність теорем, що визначають умови прямування середніх арифметичних значень випадкових величин до відповідних констант при проведенні великого числа n ($n > 100...1000$) дослідів. Самими відомими серед них є теореми

Чебишова і Бернуллі.

<u>Нерівність Чебишова</u>. *Для довільної випадкової величини* X *з математичним сподіванням* $M(X)$ *і дисперсією* $D(X)$ *справедлива нерівність*

$$\boxed{P\left(|X - M(X)| \geq \varepsilon\right) \leq \frac{D(X)}{\varepsilon^2}},$$

де ε *– довільне додатне число.* Іншими словами, *абсолютне відхилення випадкової величини від її математичного сподівання не менше* ε *з імовірністю, що не перевищує відношення дисперсії цієї випадкової величини до квадрата* ε.

<u>Приклад 1</u>. Оцінити ймовірність, що випадкова величина X прийме значення за межами інтервалу $(a_x - \varepsilon, a_x + \varepsilon)$, де $\varepsilon = 3\sigma_x$.

□ Оскільки $\sigma_x^2 = D(X)$, з нерівності Чебишова при $\varepsilon = 3\sigma_x$ маємо

$$P\left(|X - a_x| \geq 3\sigma_x\right) \leq \frac{\sigma_x^2}{9\sigma_x^2} = \frac{1}{9}. \quad \blacksquare$$

<u>Зауваження 1</u>. Нерівність Чебишова дає тільки верхню оцінку, що не залежить від закону розподілу, для ймовірності абсолютного відхилення $|X - a_x|$.

Послідовність випадкових величин X_n, $n = 1, 2, \ldots$ збігається за ймовірністю до випадкової величини X, якщо ймовірність того, що X_n відрізняється від X на довільне скінченне число прямує до нуля при $n \to \infty$:

$\boxed{X_n \underset{n\to\infty}{\overset{P}{\longrightarrow}} X}$, якщо $\lim\limits_{n\to\infty} P\left(|X_n - X| > \varepsilon\right) = 0$ для довільного $\varepsilon > 0$.

<u>Теорема Чебишова</u>. *Нехай проведено* n *однакових незалежних випробувань, у яких випадкова величина* X *прийняла відповідно значення* x_1, x_2, …, x_n. *При необмеженому збільшенні числа* n *випробувань середнє арифметичне* $\bar{x} = (1/n)\sum_{i=1}^{n} x_i$ *значень випадкової величини* X *збігається за ймовірністю до її математичного сподівання*

$$\boxed{\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \underset{n\to\infty}{\overset{P}{\to}} M(X)},$$

тобто *для довільного додатного* ε *справджується рівність*

$$\boxed{\lim_{n\to\infty} P\left(|\overline{x} - a_x| < \varepsilon\right) = 1}.$$

<u>Зауваження 2</u>. Хоча співвідношення $|\overline{x} - a_x| < \varepsilon$ не є достовірним, проте за теоремою Чебишова якщо число випробувань n досить велике, то ймовірність його виконання близька до 1, наприклад, 0,99 чи 0,999, що означає його ***практичну вірогідність***. Тобто, невідоме *математичне сподівання випадкової величини* X *практично можна замінити середнім арифметичним значенням* $\boxed{a_x \approx \overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i}$, отриманим за результатами досить великої кількості випробувань. При цьому, чим більше n, тим з більшою ймовірністю, що наближається до 1, можна очікувати, що абсолютна похибка $|\overline{x} - a_x|$ такої заміни не перевищить задану величину ε.

Наприклад, нехай проводиться серія з n вимірювань деякого показника x, причому: а) результат кожного вимірювання не залежить від результатів інших, тобто всі результати X_1, X_2, …, X_n є попарно незалежними випадковими величинами; б) вимірювання проводяться без систематичних похибок (їх математичні сподівання рівні між собою і дорівнюють істинному значенню a вимірюваної величини x); в) забезпечена певна точність вимірювань, тобто дисперсії випадкових величин X_1, X_2, …, X_n рівномірно обмежені. Тоді при достатньо великій кількості n вимірювань їх середнє арифметичне виявиться скільки завгодно близьким до істинного значення a величини x.

<u>Теорема Бернуллі</u>. *Нехай проведено* n *однакових незалежних випробувань, у кожному з яких можлива поява події* A *з імовірністю* $P(A)$, *і в результаті подія* A *відбулася* m *раз. Тоді при необмеженому зростанні числа випробувань* n ($n \to \infty$) *відносна частота* $W_n(A) = m/n$ *появи події* A *збігається за ймовірністю до її ймовірності* $P(A)$:

$$\boxed{W_n(A) \underset{n\to\infty}{\overset{P}{\to}} \mathrm{P}(A)}.$$

Іншими словами, *при* $n \to \infty$ *відносна частота випадкової події* A *з імовірністю, що прямує до 1, необмежено зближується з імовірністю цієї події* A:

$$\boxed{\lim_{n\to\infty} \mathrm{P}\left(\left|W_n(A) - \mathrm{P}(A)\right| < \varepsilon\right) = 1},$$

де ε *– довільне (як завгодно мале) фіксоване додатне число.*

Відомо, що на практиці дуже поширеним є нормальний розподіл. Пояснення цього феномену дає

центральна гранична теорема (у формі Ляпунова). *Нехай* X_1, X_2, …, X_n *– незалежні випадкові величини з приблизно однаковими дисперсіями* $D(X_i) \approx D$, $i = 1, 2, \ldots, n$. *Тоді при необмеженому зростанні* n ($n \to \infty$) *закон розподілу їх суми* $Y = \sum_{i=1}^{n} X_i$ *необмежено наближається до нормального закону*

$$F(y) = \frac{1}{\sigma_y \sqrt{2\pi}} \int_{-\infty}^{y} e^{-(u-a_y)^2 / (2\sigma_y^2)} du$$

з параметрами $a_y = \sum_{i=1}^{n} M(X_i)$ *і* $\sigma_y = \sqrt{\sum_{i=1}^{n} D(X_i)}$.

Зауваження 3. Вимога $D(X_i) \approx D$, $i = 1, 2, \ldots, n$ означає, що ні один з доданків не носить домінуючого характеру (вплив всіх X_i, $i = 1, 2, \ldots, n$ на суму Y приблизно однаковий). Таким чином, нормальний розподіл виникає тоді, коли підсумовується багато незалежних (чи слабо залежних) випадкових величин, рівень впливу кожної з яких на розсіювання суми має однаковий (рівномірно малий) порядок. На практиці така ситуація спостерігається досить часто, оскільки у випадку більш суттєвого впливу деякої величини X_i на суму Y звичайно приймаються спеціальні заходи для усунення відповідної головної причини розсіювання Y.

Зауваження 4. У реальних обставинах при підсумовуванні величин X_i, $i=1,2,...,n$ з однаковим законом розподілу, що характеризується математичним сподіванням $M(X_i)=a_x$, $i=1,2,...,n$ і дисперсією $D(X_i)=\sigma_x^2$ X_i, $i=1,2,...,n$, закон розподілу їх суми $Y=\sum_{i=1}^{n} X_i$ можна вважати приблизно нормальним за умови $n>20\,...\,50$. При цьому $a_y=n\,a_x$ і $\sigma_y=\sigma_x\sqrt{n}$, .

<u>Приклад 2</u>. Незалежні випадкові величини X_i, , $i=1,2,...,n$ $n=100$ розподілені рівномірно на відрізку $[0;1]$:

$$F(x)=\begin{cases} 0, & x<0; \\ x, & 0\le x\le 1; \\ 1, & x>1. \end{cases}$$

Знайти наближено інтегральну функцію розподілу $F(y)$ їх суми $Y=\sum_{i=1}^{n} X_i$, а також ймовірність того, що $45<Y<70$.

□ Оскільки $n=100>50$, то випадкова величина Y має приблизно нормальний розподіл $F(y)=\dfrac{1}{\sigma_y\sqrt{2\pi}}\int\limits_{-\infty}^{y} e^{-(u-a_y)^2/(2\sigma_y^2)}\,du$ з параметрами $a_y=n\,a_x$ і $\sigma_y=\sigma_x\sqrt{n}$.

Для рівномірного розподілу дістанемо:

$$f(x)=F'(x)=\begin{cases} 0, & x<0; \\ 1, & 0\le x\le 1; \\ 0, & x>1; \end{cases} \quad m_x=M(X)=\int\limits_{-\infty}^{+\infty} x\,f(x)\,dx=$$

$$=\int\limits_{0}^{1} x\cdot 1dx=x^2/2)\Big|_0^1=1/2\,;\ \ \sigma_x^2=D(X)=\int\limits_{-\infty}^{+\infty}(x-m_x)^2\,f(x)\,dx=$$

$$=\int\limits_{0}^{1}(x-1/2)^2\cdot 1dx=(x-1/2)^3/3\Big|_0^1=1/12\,.$$

Тоді

$$a_y = n\,a_x = 100\cdot 1/2 = 50\,;\;\; \sigma_y = \sigma_x\sqrt{n} = \sqrt{1/12}\cdot\sqrt{100} = 5\sqrt{3}\big/3\,.$$

Отже, $F(y) = \dfrac{3}{5\sqrt{6\pi}}\displaystyle\int_{-\infty}^{y} e^{-3(u-50)^2/50}\,du$.

Далі за формулою

$$P(\alpha < X < \beta) = \Phi\big((\beta - a_x)/\sigma_x\big) - \Phi\big((\alpha - a_x)/\sigma_x\big)$$

знайдемо ймовірність влучення випадкової величини Y в заданий діапазон $(\alpha;\beta) = (45;70)$, $\alpha < \beta$:

$$P(45 < Y < 70) \approx \Phi\left(\frac{70-50}{5\sqrt{3}/3}\right) - \Phi\left(\frac{45-50}{5\sqrt{3}/3}\right) = \Phi(4\sqrt{3}) -$$

$$-\Phi(-\sqrt{3}) = \Phi(4\sqrt{3}) + \Phi(\sqrt{3}) \approx \Phi(6{,}928) + \Phi(1{,}732) \approx$$

$$\approx 0{,}5 + 0{,}458 \approx 0{,}958\,.\quad \blacksquare$$

1.6. Контрольні запитання до змістового модулю “Теорія ймовірностей”

1. Дайте визначення випадкової (стохастичної) події.
2. Які події називаються: неможливими? достовірними? рівноможливими? несумісними? протилежними? Наведіть приклади.
3. Чи є протилежні події несумісними?
4. Чи є несумісні події протилежними?
5. Дайте визначення ймовірності випадкової події.
6. Як визначити ймовірність події класичним методом?
7. Що розуміють під повною групою подій? Наведіть приклади.
8. Як пов'язані між собою ймовірність і відносна частота появи події?
9. Як визначити ймовірність суми сумісних подій?
10. Наведіть приклади залежних і незалежних подій.
11. Що розуміють під умовною ймовірністю події?
12. Як визначається ймовірність добутку двох подій?
13. Наведіть формулу повної ймовірності.
14. Яке призначення формули Байєса?
15. Запишіть формулу Байєса. У чому її прикладне значення?

16. Який зв'язок між формулою Байєса і формулою повної ймовірності?
17. Які комбінаторні з'єднання розрізнюють у комбінаториці?
18. Дайте означення комбінаторного поняття “перестановка з n елементів”.
19. Яким способом обчислюється загальна кількість перестановок з n елементів?
20. Дайте означення комбінаторного поняття “розміщення з n елементів по m”.
21. Яким способом обчислюється загальна кількість розміщень з n елементів по m?
22. Дайте визначення комбінаторного поняття “сполучення з n елементів по m”.
23. Яким способом обчислюється загальна кількість сполучень з n елементів по m?
24. В яких випадках для визначення ймовірності застосовується формула Бернуллі?
25. Дайте визначення найімовірнішого числа появ події. Як його обчислити у випадку серії експериментів за схемою Бернуллі?
26. Наведіть локальну та інтегральну формули Лапласа.
27. В яких випадках замість формули Бернуллі застосовуються локальна й інтегральна формули Лапласа?
28. В яких випадках замість формули Бернуллі використовується формула Пуассона?
29. У випадку серії експериментів за схемою Бернуллі за якою формулою можна наближено оцінити ймовірність того, що відхилення відносної частоти $W_n(A)$ від ймовірності p за модулем не перевищує заданого додатного числа ε?
30. Чи можна стверджувати, що функція Гаусса є симетричною відносно осі ординат?
31. Чи можна стверджувати, що функція Лапласа є центрально симетричною відносно початку системи координат?
32. Дайте означення випадкової величини.
33. Яка випадкова величина називається дискретною? Наведіть приклади.
34. Яка випадкова величина називається неперервною? Наведіть приклади.

35. В яких формах може бути поданий закон розподілу випадкової величини?
36. Чи може функція розподілу бути: більшою одиниці? Від’ємною? Спадною?
37. Чому дорівнює ймовірність конкретного значення неперервної випадкової величини?
38. Що розуміють під щільністю розподілу неперервної випадкової величини?
39. Перелічіть властивості щільності розподілу.
40. Як, виходячи з ряду розподілу, знайти значення функції розподілу дискретної випадкової величини?
41. Як знайти ймовірність влучення випадкової величини в заданий діапазон, якщо відома функція розподілу? Щільність розподілу?
42. Дайте геометричну інтерпретацію ймовірності влучення неперервної випадкової величини на заданий проміжок.
43. Назвіть основні числові характеристики випадкових величин.
44. За якою формулою знаходиться математичне сподівання дискретної випадкової величини? Неперервної випадкової величини?
45. Чи є математичне сподівання випадковою величиною?
46. Чи є дисперсія випадковою величиною?
47. Як математичне сподівання і дисперсія характеризують випадкову величину?
48. Чим зручніше застосування середнього квадратичного відхилення для характеристики розсіювання замість дисперсії?
49. У чому полягає статистичний зміст параметрів a і σ нормального закону розподілу?
50. Як змінюється графік нормального закону зі зміною середнього квадратичного відхилення?
51. Дайте означення моди і медіани випадкової величини. Як мода і медіана характеризують випадкову величину?
52. Для симетричного унімодального закону розподілу випадкової величини значення математичного сподівання, моди і медіани збігаються. Чи справедливе обернене твердження?
53. Що характеризує та як визначається коефіцієнт асиметрії?
54. Чому дорівнює коефіцієнт асиметрії нормально розподіленої випадкової величини?

55. Як визначається ексцес (коефіцієнт гостровершинності) і що він характеризує?
56. Чому дорівнює ексцес нормально розподіленої випадкової величини?
57. За якими формулами знаходять математичне сподівання і дисперсію випадкової величини, що має біноміальний розподіл?
58. За якими формулами знаходять математичне сподівання і дисперсію випадкової величини, що має розподіл Пуассона?
59. Наведіть формули для обчислення математичного сподівання і дисперсії рівномірно розподіленої випадкової величини.
60. За якими формулами знаходять математичне сподівання і дисперсію випадкової величини, що має експоненціальний (показниковий) розподіл?
61. Чому дорівнює щільність розподілу випадкової величини, що описується нормальним законом?
62. Як визначити ймовірність влучення нормально розподіленої випадкової величини в заданий діапазон?
63. Як знайти ймовірність того, що відхилення нормально розподіленої випадкової величини X від її математичного сподівання $x = a$ за модулем менше заданого додатного числа δ?
64. Що таке багатомірна випадкова величина?
65. Що таке функція сумісного розподілу двох випадкових величин? Перелічіть її властивості.
66. Як визначаються математичне сподівання і дисперсія системи двох випадкових величин?
67. Як визначаються умовні математичні сподівання у випадку системи двох випадкових величин X і Y?
68. Що характеризує кореляційний момент системи двох випадкових величин?
69. Для чого використовується коефіцієнт кореляції?
70. Поясніть значення термінів: регресія, рівняння регресії, лінія регресії.
71. Що розуміють під законом великих чисел? Яка його роль у теорії ймовірностей?
72. Наведіть нерівність Чебишова і поясніть її зміст.
73. Сформулюйте теорему Чебишова і поясніть її практичний зміст.
74. Сформулюйте теорему Бернуллі і поясніть її практичний зміст.
75. Що стверджує центральна гранична теорема?

Змістовий модуль 2.

МАТЕМАТИЧНА СТАТИСТИКА

2.1. Основні задачі математичної статистики. Генеральна та вибіркова статистичні сукупності. Варіаційний ряд

Математична статистика, спираючись на методи теорії ймовірностей, займається встановленням закономірностей, яким підкоряються масові однорідні випадкові явища, на основі збору, систематизації і обробки статистичних даних, одержаних у результаті спостережень, з метою формування обґрунтованих наукових і практичних висновків в умовах стохастичної невизначеності.

Предмет математичної статистики складають методи реєстрації, опису й аналізу статистичних експериментальних даних, які можна подати як сукупність значень одно- чи багатовимірної випадкової величини, що дозволяє застосовувати апарат теорії ймовірностей. У свою чергу, математична статистика служить основою для створення методів обробки й аналізу статистичного матеріалу в різних конкретних сферах людської діяльності.

Основні задачі математичної статистики:

1) визначення способів збору, групування й опису статистичних даних;

2) розробка методів аналізу одержаних даних у залежності від мети дослідження:

а) оцінка невідомої ймовірності події; оцінка невідомої функції розподілу; оцінка параметрів розподілу відомого вигляду; оцінка залежності між випадковими величинами і т.п.;

б) перевірка правдоподібності статистичних гіпотез про вигляд невідомого розподілу чи про значення параметрів відомого розподілу.

Множину всіх статистично однорідних об’єктів, що досліджується відносно деякого якісного чи кількісного показника – випадкової величини X, називають ***генеральною сукупністю***. Число N всіх об’єктів, що утворюють генеральну сукупність, називають її ***об’ємом***.

<u>Зауваження 1</u>. Об’єм генеральної сукупності може бути нескінченним.

Часто суцільне дослідження всіх об'єктів генеральної сукупності неприйнятне, оскільки фізично складне, економічно затратне чи викликає їх пошкодження. У таких випадках з генеральної сукупності відбирають для дослідження лише деяку частину об'єктів, а одержані висновки поширюють на всю генеральну сукупність.

Вибірковою сукупністю або просто ***вибіркою*** називають підмножину об'єктів генеральної сукупності. Число n всіх об'єктів, що утворюють вибірку, називають її ***об'ємом***.

Метод статистичного дослідження, за яким на базі вивчення вибірки робиться висновок про всю генеральну сукупність, називають ***вибірковим методом***. Він служить одним з основних підходів до вивчення випадкових явищ у математичній статистиці.

Для достовірності суджень за даними вибірки про всю генеральну сукупність необхідно, щоб вибірка правильно відображала властивості генеральної сукупності, тобто була ***репрезентативною*** (***представницькою***). Виходячи з закону великих чисел, для цього вибірка повинна бути ***випадковою***, тобто складатися з випадково відібраних об'єктів генеральної сукупності.

Випадковий відбір передбачає, що кожний об'єкт генеральної сукупності має однакову ймовірність потрапити у вибірку.

Вибірка може бути ***повторною***, коли відібраний об'єкт (перед відбором наступного) повертається в генеральну сукупність, і ***безповторною***, коли відібраний об'єкт не повертається в генеральну сукупність.

Застосовують різні ***способи відбору***:

1) ***простий відбір*** – випадкове вилучення об'єктів генеральної сукупності з поверненням чи ні;

2) ***типовий відбір***, коли об'єкти вилучаються не з усієї генеральної сукупності, а лише з її "типової" частини;

3) ***серійний відбір*** – об'єкти вибирають з генеральної сукупності не по одному, а серіями;

4) ***механічний відбір*** – генеральна сукупність "механічно" ділиться на стільки частин, скільки об'єктів повинно ввійти до вибірки, і з кожної частини вилучається один об'єкт.

<u>Зауваження 2</u>. Якщо об'єм генеральної сукупності досить великий, то відмінність між повторною і безповторною вибіркою зникає.

Одержані за вибіркою статистичні дані звичайно утворюють неупорядковану множину чисел x_i, $i=\overline{1,n}$ – значень випадкової величини X, яку називають ***первинною статистичною сукупністю***. Цей безлад ускладнює виявлення закономірностей змінювання (***варіювання***) статистичних даних. Тому застосовують операцію ***ранжування***, при якій отримані значення випадкової величини X розміщують у порядку зростання.

Після ранжування наявні значення випадкової величини X групують так, щоб у кожній окремій групі значення випадкової величини було однаковим і відрізнялося від значень в усіх інших групах. Кожне таке значення називають ***варіантою***. Варіанта позначається малою літерою латинського алфавіту з індексом, що відповідає порядковому номеру групи: x_i, $i=\overline{1,k}$, де k – загальне число всіх варіант, $k \le n$.

Змінювання значень варіант називають ***варіюванням***. При ***дискретному*** варіюванні окремі значення варіант відрізняються одне від іншого на ту чи іншу скінченну величину. При ***неперервному*** варіюванні окремі значення варіант можуть відрізнятися одне від іншого на як завгодно малу величину.

Послідовність варіант, розміщених у зростаючому порядку, називають ***варіаційним рядом***: x_1, x_2, …, x_i, …, x_k, де $x_i < x_{i+1}$, $i=1,2,...,k-1$.

Різницю $\boxed{R_B = x_{\max} - x_{\min}}$, де $x_{\min} = x_1$ – найменша з варіант, а $x_{\max} = x_k$ – найбільша з варіант, називають ***розмахом*** варіаційного ряду. Відповідний відрізок $[x_{\min}; x_{\max}]$ називають ***діапазоном значень*** ряду.

Приклад. Дана вибірка $\{2,4,7,3,9,1,1,9,3,\ 2,3,7,3,7\}$. Проранжувати її в порядку зростання. Знайти розмах. Записати відповідний дискретний варіаційний ряд.

□ Проведемо ранжування: $\{1,1,2,\ 2,3,\ 3,3,3,4,7,7,7,9,9\}$.

Визначимо розмах $R_B = x_{\max} - x_{\min} = 9-1=8$.

Далі запишемо варіаційний ряд: $1,2,3,4,7,9$, $k=6$. ■

2.2. Статистичний ряд. Оцінка закону розподілу

Число n_i, що показує, скільки разів зустрічається в сукупності статистичних даних відповідна варіанта x_i (тобто об'єм i-ї групи), називають ***частотою*** цієї варіанти. Сума всіх відносних частот дорівнює об'єму вибірки: $\boxed{\sum_{i=1}^{k} n_i = n}$.

Відношення $\boxed{W_i = n_i / n}$ частоти n_i даної варіанти x_i до загальної суми $\sum_{i=1}^{k} n_i = n$ частот всіх варіант (об'єму вибірки) називають ***відносною частотою*** i-ї варіанти. Сума всіх відносних частот дорівнює одиниці: $\boxed{\sum_{i=1}^{k} W_i = 1}$.

Дискретним статистичним рядом називають проранжовану в порядку зростання послідовність усіх варіант x_1, x_2, …, x_i, …, x_k з відповідно вказаними частотами n_1, n_2, …, n_i, …, n_k або відносними частотами W_1, W_2 …, W_i, …, W_k. Дискретний статистичний ряд зручно записувати у вигляді наступної таблиці.

i	1	2	…	k	
x_i	x_1	x_2	…	x_k	$\sum_{i=1}^{k}$
n_i	n_1	n_2	…	n_k	n
W_i	W_1	W_2	…	W_k	1

<u>Приклад</u>. Для варіаційного ряду, одержаного в прикладі з попереднього пункту, скласти дискретний статистичний ряд.

□ Проведемо обчислення частот і відносних частот. Дістанемо статистичний ряд:

i	1	2	3	4	5	6	
x_i	1	2	3	4	7	9	$\sum_{i=1}^{6}$
n_i	2	2	4	1	3	2	14
W_i	1/7	1/7	2/7	1/14	3/14	1/7	1

■

Якщо випадкова величина X є неперервною (чи дискретною з великим числом різних значень), то замість дискретного часто складають ***інтервальний статистичний ряд***. Для його побудови необхідно:

1) Увесь діапазон $[x_{\min};x_{\max}]=[x_1;x_k]$ значень варіант розділити на m елементарних проміжків $[a_0;a_1)$, $[a_1;a_2)$, …, $[a_{m-2};a_{m-1})$, $[a_{m-1};a_m]$ з довжинами $h_i=a_i-a_{i-1}$, $i=\overline{1,m}$ так, що кінець попереднього служить початком наступного, причому $a_0=x_1$ і $a_m=x_k$.

2) Знайти частоти n_i і відносні частоти $W_i=n_i/n$ попадання значень випадкової величини X в i-й частинний інтервал з кінцями a_{i-1} і a_i, $i=\overline{1,m}$.

3) Обчислити значення ***емпіричної щільності розподілу*** $\boxed{f^*(x)=W_i/h_i\,,\quad x\in[a_{i-1};a_i)}$ – щільності відносної частоти W_i, $i=\overline{1,m}$.

4) Скласти відповідну таблицю

i	1	2	…	$m-1$	m	–
x_i	$[a_0;a_1)$	$[a_1;a_2)$	…	$[a_{m-2};a_{m-1})$	$[a_{m-1};a_m]$	$\sum_{i=1}^{m}$
n_i	n_1	n_2	…	n_{m-1}	n_m	n
W_i	W_1	W_2	…	W_{m-1}	W_m	1
W_i/h_i	W_1/h_1	W_2/h_2	…	W_{m-1}/h_{m-1}	W_m/h_m	–

Зауваження 1. Надалі обмежимося розглядом тільки розбиття на елементарні проміжки однакової довжини $h_i=h=(x_k-x_1)/m$, $i=\overline{1,m}$. Тоді $a_0=x_1$ і $a_i=a_{i-1}+h=x_1+ih$, $i=\overline{1,m}$.

Зауваження 2. При групуванні даних з метою побудови інтервального статистичного ряду виникає питання раціонального вибору числа m елементарних інтервалів. Для цього рекомендується наближені співвідношення:

$$\boxed{m\approx \operatorname{int}(n^{1/2}),\quad n\le 100} \text{ або } \boxed{m\approx \operatorname{int}(1+3{,}322\lg n),\quad n>100},$$

де $\mathrm{int}(x)$ – ціла частина дійсного числа x. Бажано вибрати m так, щоб n націло ділилось на m.

Для дискретної випадкової величини X відносна частота W_i служить статистичною оцінкою за вибіркою ймовірності появи відповідної варіанти x_i. Тому дискретний статистичний ряд з наведеними значеннями відносної частоти W_i служить наближеним поданням ряду розподілу величини X.

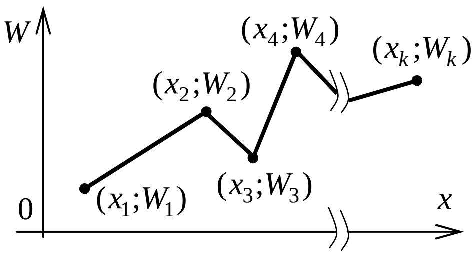


Рис. 34

У прямокутній системі координат Oxy ламану з вершинами $(x_i;W_i)$, $i=\overline{1,k}$ називають ***полігоном відносних частот*** (рис. 34). Він є статистичним емпіричним аналогом многокутника розподілу.

Аналогічно, для неперервної випадкової величини X інтервальний статистичний ряд з наведеними значеннями емпіричної щільності $f^*(x)=W_i/h_i$, $x\in[a_{i-1};a_i)$, $i=\overline{1,m}$ служить наближеним поданням щільності розподілу $f(x)$ величини X. У прямокутній системі координат Oxy східчасту фігуру, складену з m елементарних прямокутників, у кожного з яких основою служить відповідний частинний інтервал з кінцями a_{i-1} і a_i, а висота дорівнює відповідному значенню $\boxed{f^*(x)=W_i/h}$ щільності відносної частоти W_i, $i=\overline{1,m}$, називають ***гістограмою відносних частот*** (рис. 35). Вона є статистичним емпіричним аналогом графіка щільності розподілу.

Наближеною оцінкою інтегральної функції розподілу служить ***емпірична функція розподілу*** $F^*(x)$, що за вибіркою для кожного значення x визначає відносну частоту події $X<x$. Щоб знайти $F^*(x)$, треба підрахувати число варіант $n(x)$, у яких випадкова величина X прийняла значення, менші x. Тоді $\boxed{F^*(x)=n(x)/n}$.

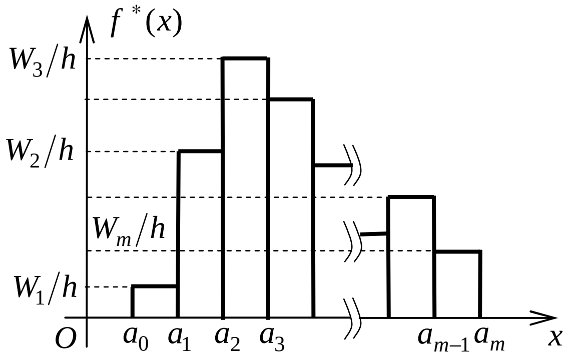


Рис. 35

За теоремою Бернуллі зі збільшенням об'єму вибірки n при будь-якому x відносна частота події $X < x$ збігається за ймовірністю до ймовірності цієї події. Отже, при необмеженому зростанні n ($n \to \infty$) емпірична функція розподілу $F^*(x)$ збігається за ймовірністю до істинної функції розподілу $F(x)$ випадкової величини X.

У прямокутній системі координат Oxy графік емпіричної функції розподілу $F^*(x)$ називають ***кумулятивною кривою*** (***кумулятою***).

У випадку дискретного статистичного ряду величина $F^*(x_i)$ для варіанти x_i є накопиченою відносною частотою, що утворюється підсумовуванням відносних частот всіх варіант, що передують даній: $\boxed{F^*(x_i) = \sum_{j=1}^{i-1} W_j}$. Кумулятою служить кусково-стала східчаста лінія, що задається співвідношенням:

$$\boxed{F^*(x) = \begin{cases} 0, \quad x \in (-\infty; x_1]; \\ \sum_{j=1}^{i-1} W_j, \ x \in (x_{i-1}; x_i], \ i = \overline{2,k}; \\ 1, \quad x \in (x_k; +\infty). \end{cases}}$$

Для будь-якої випадкової величини X – дискретної чи неперервної – кумулята є розривною кусково-сталою східчастою лінією, скінченні стрибки якої відповідають значенням варіант і за величиною дорівнюють їх відносним частотам. Якщо X – неперервна величина, то при збільшенні об'єму вибірки n число стрибків збільшується, а їх висоти зменшуються. При цьому графік емпіричної функції розподілу $F^*(x)$ необмежено наближається до неперервної кривої, що відповідає істинній функції розподілу $F(x)$.

2.3. Точкові статистичні оцінки числових характеристик випадкових величин

Кожній числовій характеристиці (параметру) θ розподілу випадкової величини X відповідає її ***статистична оцінка*** θ_B – наближене значення, одержане за результатами спостережень x_i, $i=\overline{1,n}$ за вибіркою об'єму n, тобто відповідна ***вибіркова характеристика***.

Розрізняють точкові та інтервальні статистичні оцінки.

Точкова оцінка θ_B задається одним числом як деяка функція від n одержаних за вибіркою випадкових величин x_i, $i=\overline{1,n}$. Значення θ_B змінюються випадковим чином при переході від однієї вибірки до іншої. Тобто, значення вибіркових характеристик містять, на відміну від самих числових характеристик, елемент випадковості.

Інтервальна оцінка визначається двома числами θ_{B1} і θ_{B2} – кінцями ***довірчого інтервалу*** $(\theta_{B1};\theta_{B2})$, що з заданою надійністю γ накриває істинне значення θ. ***Надійністю*** (***довірчою ймовірністю***) інтервальної оцінки називають близьку до 1 ймовірність γ (наприклад, $\gamma=0{,}9$, $\gamma=0{,}95$ або $\gamma=0{,}99$), з якою ця оцінка відповідає істинному значенню θ, тобто ймовірність події $\theta\in(\theta_{B1};\theta_{B2})$. Значення γ вибирають таким, що подію з імовірністю γ можна вважати практично достовірною. Ймовірність α протилежної події $\theta\notin(\theta_{B1};\theta_{B2})$ називають ***рівнем значущості*** інтервальної оцінки. При цьому $\alpha=1-\gamma$.

Для якісної заміни істинного значення числової характеристики θ його наближеною точковою оцінкою θ_B остання повинна відповідати певним критеріям – спроможності, незміщеності та ефективності.

Оцінка θ_B називається ***спроможною***, якщо при необмеженому зростанні n ($n \to \infty$) вона збігається за ймовірністю до істинного значення θ: $\theta_B \underset{n\to\infty}{\overset{P}{\to}} \theta$.

Спроможність – це мінімальна вимога до оцінки. Неспроможні оцінки не використовуються. Вимога спроможності оцінки страхує від появи грубих похибок при досить великих об'ємах вибірки.

Оцінка θ_B називається ***незміщеною***, якщо її математичне сподівання $M(\theta_B)$ дорівнює істинному значенню θ при довільному об'ємі вибірки: $M(\theta_B) = \theta$, $\forall n$. Тобто, незміщена оцінка θ_B не містить систематичної похибки.

Незміщена оцінка θ_B є спроможною, якщо її дисперсія $D(\theta_B)$ задовольняє умові $\lim\limits_{n\to\infty} D(\theta_B) = 0$.

Інколи оцінка θ_B є ***асимптотично незміщеною***, тобто $M(\theta_B) \underset{n\to\infty}{\overset{P}{\to}} \theta$.

Вимога незміщеності особливо суттєва при малому об'ємі вибірки, оскільки усуває можливі систематичні похибки.

Оцінка θ_B називається ***ефективною***, якщо при заданому об'ємі вибірки вона має найменшу дисперсію $D(\theta_B)$ серед всіх можливих оцінок.

Вимога ефективності використовується для вибору оцінки з найменшим розсіюванням.

<u>Оцінка математичного сподівання</u>. *Незміщеною спроможною оцінкою математичного сподівання* $M(X)$ *випадкової величини* X *служить* ***вибіркове середнє*** $\overline{x}$ – середнє арифметичне всіх вибіркових значень x_i, $i = \overline{1,n}$: $\boxed{M(X) \approx \overline{x}}$.

Якщо всі значення x_i, $i = \overline{1,n}$ різні, то $\boxed{\overline{x} = (1/n)\sum_{i=1}^{n} x_i}$.

Якщо кожне значення x_i має відповідну частоту n_i, $i=\overline{1,k}$, причому $\sum_{i=1}^{k} n_i = n$, то $\boxed{\bar{x} = (1/n)\sum_{i=1}^{k} n_i x_i}$.

Зауваження 1. Вибіркове середнє $\bar{x}$ є ефективною оцінкою для $M(X)$, якщо випадкова величина X розподілена за нормальним законом.

Оцінка дисперсії та середнього квадратичного відхилення. *Спроможною оцінкою дисперсії $D(X)$ випадкової величини X служить* ***вибіркова дисперсія*** $D_B(X)$ – середнє арифметичне всіх квадратів відхилень вибіркових значень x_i, $i=\overline{1,n}$ від їх вибіркового середнього $\bar{x}$: $\boxed{D(X) \approx D_B(X)}$.

Якщо всі значення x_i, $i=\overline{1,n}$ різні, то

$$\boxed{D_B(X) = (1/n)\sum_{i=1}^{n} (x_i - \bar{x})^2}.$$

Якщо кожне значення x_i має відповідну частоту n_i, $i=\overline{1,k}$, причому $\sum_{i=1}^{k} n_i = n$, то $\boxed{D_B(X) = (1/n)\sum_{i=1}^{k} n_i (x_i - \bar{x})^2}$.

Вибірковим середнім квадратичним відхиленням $\sigma_B(X)$ називають квадратний корінь з вибіркової дисперсії $D_B(X)$: $\boxed{\sigma_B = \sqrt{D_B}}$. $\sigma_B(X)$ *служить спроможною оцінкою середнього квадратичного відхилення* $\sigma(X)$: $\boxed{\sigma(X) \approx \sigma_B(X)}$.

Зауваження 2. Вибіркова дисперсія $D_B(X)$ є зміщеною оцінкою для $D(X)$. *Незміщеною спроможною оцінкою дисперсії* $D(X)$ *служить* ***виправлена дисперсія*** $S^2(X)$, яка обчислюється за формулою $\boxed{S^2(X) = \big(n/(n-1)\big)D_B(X)}$. При достатньо великих n виправлена і вибіркова дисперсії мало відрізняються. Відповідно *незміщеною спроможною оцінкою середнього квадратичного відхилення* $\sigma(X)$ *служить* ***виправлене середнє квадратичне відхилення*** $S(X)$: $\boxed{S(X) = \sqrt{S^2(X)}}$.

Зауваження 3. Наведемо ще дві характеристики дискретного статистичного ряду з k варіантами x_1, x_2, …, x_k, що служать статистичними оцінками відповідних числових характеристик випадкової величини X :

1) ***Мода*** Mo_B – варіанта, що має найбільшу частоту.

2) ***Медіана*** Me_B – значення випадкової величини X, що відповідає середині ряду (з врахуванням частот варіант). При цьому:

а) $Me_B = x_{(n+1)/2}$, якщо n – непарне число;

б) $Me_B = (x_{n/2} + x_{n/2+1})/2$, якщо n – парне число.

Приклад. Вибірка задана дискретним статистичним рядом

i	1	2	3	4	5	6	7	
x_i	-7	-1	2	5	7	8	10	$\sum_{i=1}^{7}$
n_i	10	30	20	10	2	2	6	80

Знайти: вибіркове середнє $\overline{x}$, вибіркову дисперсію $D_B(X)$, вибіркове середнє квадратичне відхилення $\sigma_B(X)$, виправлену дисперсію $S^2(X)$, виправлене середнє квадратичне відхилення $S(X)$, моду Mo_B, медіану Me_B.

□ Число варіант $k = 7$, об'єм вибірки $n = 80$ – парне число. Знайдемо

$$Mo_B = -1; \quad Me_B = (x_{n/2} + x_{n/2+1})/2 = (x_{80/2} + x_{80/2+1})/2 =$$
$$= (x_{40} + x_{41})/2 = (-1+2)/2 = 0{,}5.$$

Проведемо обчислення інших характеристик:

$$\overline{x} = (1/n)\sum_{i=1}^{k} n_i\, x_i = (10\cdot(-7) + 30\cdot(-1) + 20\cdot 2 + 10\cdot 5 +$$

$$+ 2\cdot 7 + 2\cdot 8 + 6\cdot 10)/80 = 1; \quad D_B(X) = \frac{1}{n}\sum_{i=1}^{k} n_i(x_i - \overline{x})^2 =$$

$$= (10\cdot(-7-1)^2 + 30\cdot(-1-1)^2 + 20\cdot(2-1)^2 + 10\cdot(5-1)^2 +$$
$$+ 2\cdot(7-1)^2 + 2\cdot(8-1)^2 + 6\cdot(10-1)^2)/80 = 19{,}95;$$

$$\sigma_B(X) = \sqrt{D_B(X)} = \sqrt{19{,}95} = 4{,}467\,; \quad S^2(X) = \frac{n}{n-1} D_B(X) =$$

$$= \left(80/(80-1)\right) \cdot 19{,}95 = 20{,}20\,; \quad S(X) = \sqrt{20{,}20} = 4{,}495\,. \ \blacksquare$$

2.4. Інтервальні статистичні оцінки числових характеристик випадкових величин

Нехай для числової характеристики (параметра) θ розподілу випадкової величини X за результатами спостережень x_i, $i = \overline{1,n}$ за вибіркою об'єму n одержана незміщена точкова оцінка θ_B. Оцінимо можливу абсолютну похибку $\Delta = |\,\theta_B - \theta\,|$, що виникає при наближеній заміні істинного значення θ вибірковим θ_B.

Задамо достатній рівень довірчої ймовірності γ і знайдемо таке значення $\varepsilon > 0$, для якого виконується рівність $P(\Delta < \varepsilon) = \gamma$. Тоді великі абсолютні похибки $\Delta > \varepsilon$ будуть з'являтися тільки з малою ймовірністю $\alpha = 1 - \gamma$. Враховуючи, що $\Delta = |\,\theta_B - \theta\,|$, і розкриваючи модуль, рівність $P(\Delta < \varepsilon) = \gamma$ можна подати у вигляді $P(\theta_B - \varepsilon < \theta < \theta_B + \varepsilon) = \gamma$. Це означає, що з імовірністю γ істинне значення θ потрапляє в довірчий інтервал $(\theta_B - \varepsilon;\, \theta_B + \varepsilon)$. Оскільки цей інтервал визначається за вибіркою і є випадковим, даний факт краще сформулювати так: *довірчий інтервал* $(\theta_B - \varepsilon;\, \theta_B + \varepsilon)$ *з імовірністю* γ *накриває істинне значення* θ.

Зауваження 1. Для побудови довірчого інтервалу необхідно розв'язати відносно ε рівняння $P(\Delta < \varepsilon) = \gamma$. А це потребує знання закону розподілу випадкової величини θ_B, який, в свою чергу, залежить від невідомого закону розподілу величини X, отже від його оцінюваних параметрів, у тому числі істинного значення θ. Практичне розв'язання цієї проблеми базується на використанні властивостей відповідної оцінки і певних припущень про характер розподілу величини X, що приводить до спеціальних розподілів (статистик), що відповідають тим чи іншим оцінкам.

Зауваження 2. Довірча ймовірність γ характеризує надійність інтервальної оцінки: чим більше значення γ, тим рідше довірчий інтервал не накриватиме істинне значення параметра. Довжина довірчого інтервалу 2ε характеризує якість цієї оцінки. Об'єм n використаних для одержання оцінки статистичних даних характеризує її вартість. При фіксованій вартості підвищення якості інтервальної оцінки призводить до зниження її надійності. Підвищення надійності оцінки при збереженні якості вимагає збільшення вартості.

Зауваження 3. Інколи застосовують односторонні довірчі інтервали, межі яких визначаються з умови $P(\theta_B - \varepsilon < \theta) = \gamma$ або $P(\theta < \theta_B + \varepsilon) = \gamma$.

Довірчий інтервал для математичного сподівання та дисперсії.

1) Нехай закон розподілу випадкової величини X невідомий. Відповідно до центральної граничної теореми, закон розподілу вибіркової середньої $\bar{x}$ – точкової оцінки математичного сподівання $a = M(X)$ – близький до нормального з параметрами $a_B = a$ і $\sigma_B = \sigma(X)/\sqrt{n}$. Довірчі інтервали з надійністю γ для математичного сподівання $a = M(X)$ і дисперсії $D(X) = \sigma^2$ мають вигляд:

$$\boxed{\bar{x} - S(X)\,z_{кр}\big/\sqrt{n} < a < \bar{x} + S(X)\,z_{кр}\big/\sqrt{n}};$$

$$\boxed{S^2(X)\left(1 - z_{кр}\sqrt{2/(n-1)}\right) < D(X) < S^2(X)\left(1 + z_{кр}\sqrt{2/(n-1)}\right)},$$

де $z_{кр}$ – відповідне значення аргументу функції Лапласа, що визначається зі співвідношення $\boxed{\Phi(z_{кр}) = \gamma/2}$. (Величина $z_{кр}$ є квантилем порядку $(1+\gamma)/2$ стандартного нормального розподілу $F^*(x) = 1/2 + \Phi(x)$ з параметрами $a = 0$ і $\sigma = 1$).

2) Нехай випадкова величина X розподілена за нормальним законом з невідомими параметрами a і σ. Довірчий інтервал з надійністю γ для математичного сподівання $a = M(X)$ має вигляд:

$$\boxed{\bar{x} - S(X)t_{кр}\big/\sqrt{n} < a < \bar{x} + S(X)t_{кр}\big/\sqrt{n}},$$

де $\boxed{t_{кр} = t_{\alpha,k}}$ – відповідне значення, взяте з таблиці критичних точок ***розподілу Стьюдента*** з $k = n-1$ ступенями свободи для випадку двосторонньої критичної області; $\alpha = 1 - \gamma$ – рівень значущості.

Довірчий інтервал з надійністю γ для дисперсії $D(X) = \sigma^2$ має вигляд:

$$\boxed{(n-1)S^2(X) \Big/ \chi^2_{кр2} < D(X) < (n-1)S^2(X) \Big/ \chi^2_{кр1}},$$

де $\boxed{\chi^2_{кр1} = \chi^2_{(1+\gamma)/2,k}}$ і $\boxed{\chi^2_{кр2} = \chi^2_{(1-\gamma)/2,k}}$ – відповідні значення, взяті з таблиці критичних точок χ^2 ***-розподілу*** з $k = n - 1$ ступенями свободи.

<u>Приклад 1</u>. Випадкова величина X розподілена за нормальним законом з невідомими параметрами a і σ. За вибіркою об'єму $n = 25$ знайдено її вибіркове середнє $\bar{x} = -1{,}74$ і виправлене середнє квадратичне відхилення $S(X) = 0{,}83$. Визначити довірчі інтервали з надійністю $\gamma = 0{,}95$ для математичного сподівання $a = M(X)$ і дисперсії $D(X)$.

□ Маємо: $k = n - 1 = 25 - 1 = 24$ – число ступенів свободи; $\alpha = 1 - \gamma = 1 - 0{,}95 = 0{,}05$ – рівень значущості.

Користуючись таблицею критичних точок розподілу Стьюдента з k ступенями свободи для випадку двосторонньої критичної області, знаходимо: $t_{кр} = t_{\alpha,k} = t_{0{,}05;24} = 2{,}06$. Далі проводимо обчислення довірчого інтервалу для математичного сподівання:

$$-1{,}74 - 0{,}83 \cdot 2{,}06 \Big/ \sqrt{25} < a < -1{,}74 + 0{,}83 \cdot 2{,}06 \Big/ \sqrt{25};$$

$$-2{,}08 < a < -1{,}40.$$

Користуючись таблицею критичних точок χ^2-розподілу з k ступенями свободи знаходимо:

$$\chi^2_{кр1} = \chi^2_{(1+\gamma)/2,k} = \chi^2_{(1+0{,}95)/2,24} = \chi^2_{0{,}975;24} = 12{,}4;$$

$$\chi^2_{кр2} = \chi^2_{(1-\gamma)/2,k} = \chi^2_{(1-0{,}95)/2,24} = \chi^2_{0{,}025;24} = 39{,}4.$$

Далі проводимо обчислення довірчого інтервалу для дисперсії:

$$(25-1) \cdot 0{,}83^2 \Big/ 39{,}4 < D(X) < (25-1) \cdot 0{,}83^2 \Big/ 12{,}4;$$

$$0{,}42 < D(X) < 1{,}33. \qquad ■$$

Зауваження 4. Відомо, що χ^2-розподіл при $n \to \infty$ необмежено наближається до нормального. Тому при достатньо великому об'ємі вибірки $(n \geq 50)$ довірчий інтервал для дисперсії можна знайти так:

$$S^2(X) \Big/ \left(1 + z_{кр} \big/ \sqrt{2n}\right)^2 < D(X) < S^2(X) \Big/ \left(1 - z_{кр} \big/ \sqrt{2n}\right)^2 .$$

3) Нехай випадкова величина X розподілена за нормальним законом з відомим параметром σ (середнім квадратичним відхиленням). Тоді довірчий інтервал з надійністю γ для математичного сподівання $a = M(X)$ має вигляд:

$$\boxed{\bar{x} - \sigma z_{кр} \big/ \sqrt{n} < a < \bar{x} + \sigma z_{кр} \big/ \sqrt{n}}.$$

Приклад 2. Випадкова величина X розподілена за нормальним законом з відомим середнім квадратичним відхиленням $\sigma = 1{,}37$. За вибіркою об'єму $n = 64$ знайдено її вибіркове середнє $\bar{x} = 5{,}82$. Визначити довірчий інтервал з надійністю $\gamma = 0{,}99$ для математичного сподівання $a = M(X)$.

□ Зі співвідношення $\Phi(z_{кр}) = \gamma/2 = 0{,}99/2 = 0{,}495$ за таблицею функції Лапласа знаходимо $z_{кр} = 2{,}58$. Далі проводимо обчислення довірчого інтервалу:

$$5{,}82 - 1{,}37 \cdot 2{,}58 \big/ \sqrt{64} < a < 5{,}82 + 1{,}37 \cdot 2{,}58 \big/ \sqrt{64};$$

$$5{,}38 < a < 6{,}26. \quad ■$$

2.5. Статистичне дослідження залежностей

Нехай у результаті спостережень над вибіркою об'єму n одержана неупорядкована первинна статистична сукупність значень (x_i, y_i), $i = \overline{1, n}$ двовимірної випадкової величини (X, Y). При цьому одне й те ж значення x_i або y_j, чи пара (x_i, y_j) можуть зустрічатися декілька разів. Тому первинні статистичні дані звичайно упорядковують, групують і формують так звану ***кореляційну таблицю***:

$Y \backslash X$	x_1	x_2	$\ldots$	x_i	$\ldots$	x_k	$n(y)$
y_1	q_{11}	q_{21}	$\ldots$	q_{i1}	$\ldots$	q_{k1}	n_1
y_2	q_{12}	q_{22}	$\ldots$	q_{i2}	$\ldots$	q_{k2}	n_2
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
y_j	q_{1j}	q_{2j}	$\ldots$	q_{ij}	$\ldots$	q_{kj}	n_j
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
y_l	q_{1l}	q_{2l}	$\ldots$	q_{il}	$\ldots$	q_{kl}	n_l
$m(x)$	m_1	m_2	$\ldots$	m_i	$\ldots$	m_k	n

Тут x_1, x_2, …, x_i, …, x_k – розміщені в порядку зростання різні між собою значення випадкової величини X; k – кількість різних значень X; y_1, y_2, …, y_j, …, y_l – розміщені в порядку зростання різні між собою значення випадкової величини Y; l – кількість різних значень Y; m_i – частота появи значення x_i; n_j – частота появи значення y_j; q_{ij} – частота появи пари (x_i, y_j). При цьому $\sum_{i=1}^{k} m_i = n$; $\sum_{j=1}^{l} n_j = n$; $\sum_{i,j} q_{ij} = n$.

Кореляційна таблиця є вибірковим аналогом матриці розподілу, що відображає теоретичний закон розподілу дискретної двовимірної випадкової величини (X,Y).

Статистична обробка наведених у кореляційній таблиці вибіркових даних про двовимірну випадкову величину (X,Y) включає стандартні процедури оцінки й аналізу параметрів окремих складових X і Y як одновимірних величин (відповідні підходи розглянуті вище), а також відповідні розрахунки оцінок й аналіз параметрів, притаманних тільки двовимірним випадковим величинам. Звичайно визначаються наступні оцінки числових характеристик двовимірної випадкової величини (X,Y).

<u>Оцінки математичних сподівань</u>:

$$\boxed{M(X) \approx \bar{x} = (1/n)\sum_{i=1}^{k} m_i x_i}; \quad \boxed{M(Y) \approx \bar{y} = (1/n)\sum_{j=1}^{l} n_j y_j}.$$

Оцінки дисперсій:

$$\boxed{D(X)\approx D_B(X)=\frac{1}{n}\sum_{i=1}^{k} m_i(x_i-\bar{x})^2\ ;D(X)\approx S^2(X)=\frac{n}{n-1}D_B(X)};$$

$$\boxed{D(Y)\approx D_B(Y)=\frac{1}{n}\sum_{j=1}^{l} n_j(y_j-\bar{y})^2\ ;\ D(Y)\approx S^2(Y)=\frac{n}{n-1}D_B(Y)}.$$

Оцінки умовних математичних сподівань:

$$\boxed{a_{x/y_j}=M(X/y_j)\approx\bar{x}(y_j)},\quad j=\overline{1,l}\ ;$$

$$\boxed{a_{y/x_i}=M(Y/x_i)\approx\bar{y}(x_i)},\quad i=\overline{1,k}\ ,$$

де $\bar{x}(y_j)$ і $\bar{y}(x_i)$ – ***умовні вибіркові середні***:

$$\boxed{\bar{x}(y_j)=\frac{1}{n_j}\sum_i x_i\, q_{ij}};\qquad \boxed{\bar{y}(x_i)=\frac{1}{m_i}\sum_j y_j\, q_{ij}}.$$

Оцінка коефіцієнта кореляції: $\boxed{r_{xy}\approx r_B}$,

де r_B – ***вибірковий коефіцієнт кореляції***: $\boxed{r_B=\dfrac{\overline{xy}-\bar{x}\cdot\bar{y}}{\sigma_B(X)\cdot\sigma_B(Y)}}$.

Тут $\overline{xy}$ – ***вибіркове змішане середнє***: $\boxed{\overline{xy}=(1/n)\sum_{i,j} x_i y_j\, q_{ij}}$.

Сумісне дослідження випадкових величин X і Y є предметом ***кореляційного аналізу***. Його мета – визначення форми залежності між величинами X і Y та тісноти зв'язку між ними, тобто одержання емпіричної оцінки $\bar{y}(x)$ умовного математичного сподівання $a_{y/x}$ як деякої функції від x: $\bar{y}(x)=\varphi(x,\theta_1,\theta_2,...,\theta_m)$ – вибіркової оцінки регресії Y на X (або, навпаки, регресії X на Y). Тут $\theta_1,\theta_2,...,\theta_m$ – невідомі параметри.

Для цього необхідно, по-перше, провести ***структурну ідентифікацію*** залежності $\bar{y}(x)=\varphi(x,\theta_1,\theta_2,...,\theta_m)$ – встановити клас, з якого вибирається функція $\varphi(x,\theta_1,\theta_2,...,\theta_m)$, тобто вияснити, чи є вона лінійною, квадратичною, логарифмічною, показниковою і т.д.; по-друге, провести ***параметричну ідентифікацію*** залежності $\bar{y}(x)=\varphi(x,\theta_1,\theta_2,...,\theta_m)$ – визначити значення невідомих коефіцієнтів $\theta_1,\theta_2,...,\theta_m$ функції $\varphi(x,\theta_1,\theta_2,...,\theta_m)$ з вибраного класу.

Для визначення типу залежності $\bar{y}(x)=\varphi(x,\theta_1,\theta_2,...,\theta_m)$ часто використовуються геометричні міркування: за кореляційною таблицею в прямокутній системі координат будується ***діаграма розсіювання*** (***кореляційне поле***) – множина точок (x_i,y_i), $i=\overline{1,k}$, $j=\overline{1,l}$ (рис. 36). Аналізуючи вигляд діаграми розсіювання, вибирають тип ***емпіричної*** (***вибіркової***) ***лінії регресії*** $\bar{y}(x)=\varphi(x,\theta_1,...,\theta_m)$, що повинна проходити через множину вибіркових точок так, щоб її графік найкращим чином відповідав невідомій істинній лінії регресії, тобто її значення повинні приблизно дорівнювати середнім арифметичним $\bar{y}(x)$ значень випадкової величини Y для кожного значення $X=x$.

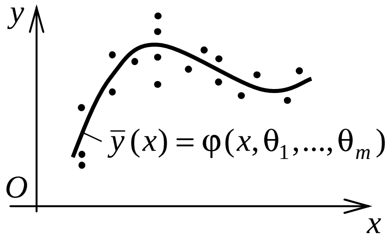


Рис. 36

Замість діаграми розсіювання можна будувати ***емпіричну*** (***вибіркову***) ***ламану регресії*** Y на X. Треба обчислити вибіркові середні $\bar{y}(x_i)$, $i=\overline{1,k}$, потім зобразити точки $(x_i,\bar{y}(x_i))$, $i=\overline{1,k}$ на координатній площині Oxy і сполучити їх послідовно відрізками прямих (рис. 37).

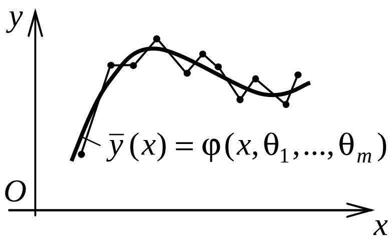


Рис. 37

Зауваження 1. У багатьох випадках тип апроксимуючої функції $\varphi(x,\theta_1,\theta_2,...,\theta_m)$ вибирається, виходячи з теоретичних міркувань, на основі аналізу предметної області, в рамках якої будується рівняння регресії.

Для визначення оцінок невідомих параметрів $\theta_1,\theta_2,...,\theta_m$, які забезпечують найкраще узгодження кривої $\bar{y}(x)=\varphi(x,\theta_1,...,\theta_m)$ з експериментальними точками, найчастіше використовується ***метод***

найменших квадратів (МНК).

Суть МНК: оптимальні оцінки коефіцієнтів $\theta_1, \theta_2, ..., \theta_m$ вибираються з умови, щоб зважена сума квадратів $s = \sum_{i=1}^{k} s_i^2 m_i$ ***нев'язок*** (відхилень) $s_i = \varphi(x_i, \theta_1, ..., \theta_m) - \bar{y}(x_i)$ між одержаним за вибіркою умовним середнім $\bar{y}(x_i)$ випадкової величини Y та його наближеним значенням $\varphi(x_i, \theta_1, ..., \theta_m)$, обчисленим за вибраною регресійною моделлю, досягала найменшого з можливих значень:

$$\sum_{i=1}^{k} \left(\varphi(x_i, \theta_1, ..., \theta_m) - \bar{y}(x_i)\right)^2 m_i \xrightarrow[\theta_1, ..., \theta_m]{} \min .$$

<u>Зауваження 2</u>. Якщо похибками значень величини X можна знехтувати й усі спостереження величини Y здійснюються з однаковою точністю та їх похибки розподілені за нормальним законом, а вибраний клас апроксимуючої функції збігається з істинним, то знайдена за МНК оптимальна залежність є найімовірнішою з усіх можливих функцій.

Нехай розміщення точок на діаграмі розсіювання нагадує пряму. Тоді природно шукану регресійну залежність вважати лінійною функцією: $\boxed{y = Ax + B}$. При цьому зважена сума квадратів усіх відхилень

$$s = \sum_{i=1}^{k} s_i^2 m_i = \sum_{i=1}^{k} \left(Ax_i + B - \bar{y}(x_i)\right)^2 m_i$$

є квадратичною функцією параметрів моделі $s = s(A, B)$. Тому вона має єдиний мінімум, для знаходження якого досить скористатися необхідними умовами екстремуму:

$$\begin{cases} \partial s / \partial A = 0; \\ \partial s / \partial B = 0; \end{cases} \qquad \begin{cases} 2\sum_{i=1}^{k} m_i x_i (Ax_i + B - \bar{y}(x_i)) = 0; \\ 2\sum_{i=1}^{k} m_i (Ax_i + B - \bar{y}(x_i)) = 0; \end{cases}$$

$$\begin{cases} \left(\sum_{i=1}^{k} x_i^2 m_i\right) A + \left(\sum_{i=1}^{k} x_i m_i\right) B = \sum_{i=1}^{k} x_i \, \bar{y}(x_i) m_i ; \\ \left(\sum_{i=1}^{k} x_i m_i\right) A + nB = \sum_{i=1}^{k} \bar{y}(x_i) m_i . \end{cases}$$

Остання система називається ***нормальною системою*** методу найменших квадратів. Розв'язуючи цю систему, знаходимо шукані оптимальні значення параметрів A і B вибіркової регресії:

$$A=\frac{n\sum_{i=1}^{k}x_i\bar{y}(x_i)m_i-\sum_{i=1}^{k}x_im_i\cdot\sum_{i=1}^{k}\bar{y}(x_i)m_i}{n\sum_{i=1}^{k}x_i^2m_i-\left(\sum_{i=1}^{k}x_im_i\right)^2};$$

$$B=\frac{\sum_{i=1}^{k}x_i^2m_i\cdot\sum_{i=1}^{k}\bar{y}(x_i)m_i-\sum_{i=1}^{k}x_im_i\cdot\sum_{i=1}^{k}x_i\bar{y}(x_i)m_i}{n\sum_{i=1}^{k}x_i^2m_i-\left(\sum_{i=1}^{k}x_im_i\right)^2}.$$

Розділивши почленно чисельник і знаменник кожної з одержаних оцінок на n^2, дістанемо їх вирази через вибіркові оцінки числових характеристик системи випадкових величин (X,Y):

$$\boxed{A=r_B\cdot\sigma_B(Y)/\sigma_B(X)};\qquad \boxed{B=\bar{y}-A\bar{x}}.$$

<u>Приклад</u>. Дана кореляційна таблиця:

$Y \backslash X$	-4	-3	1	1,5	2	$n(y)$
-1	0	0	2	5	3	10
2	0	3	6	3	2	14
7	0	8	1	0	0	9
9	6	1	0	0	0	7
$m(x)$	6	12	9	8	5	40

Обчислити: вибіркові середні $\bar{x}$ і $\bar{y}$, вибіркові дисперсії $D_B(X)$ і $D_B(Y)$, вибіркові середні квадратичні відхилення $\sigma_B(X)$ і $\sigma_B(Y)$, умовні вибіркові середні $\bar{x}(y_j)$ і $\bar{y}(x_i)$, вибіркове змішане середнє $\overline{xy}$, вибірковий коефіцієнт кореляції r_B, оптимальні значення параметрів A і B лінійної вибіркової регресії $y=Ax+B$. Записати рівняння лінійної регресії. В одній системі координат Oxy побудувати емпіричну ламану регресії Y на X і знайдену пряму регресії.

<u>Вказівка</u>. Значення шуканих величин подати з точністю до одного десяткового знака після коми.

□ Тут $n = 40$, $k = 5$, $l = 4$. Проведемо обчислення:

$$\overline{x} = (1/n)\sum_{i=1}^{k} m_i\, x_i = (1/40)(-4\cdot 6 + (-3)\cdot 12 + 1\cdot 9 + 1{,}5\cdot 8 +$$

$$+ 2\cdot 5) = -0{,}725 = -0{,}7\,; \quad \overline{y} = (1/n)\sum_{j=1}^{l} n_j\, y_j = (1/40)(-1\cdot 10 +$$

$$+ 2\cdot 14 + 7\cdot 9 + 9\cdot 7) = 3{,}6\,; \quad D_B(X) = \frac{1}{n}\sum_{i=1}^{k} m_i (x_i - \overline{x})^2 = (1/40)\times$$

$$\times((-4 + 0{,}725)^2\cdot 6 + (-3 + 0{,}725)^2\cdot 12 + (1 + 0{,}725)^2\cdot 9 +$$

$$+ (1{,}5 + 0{,}725)^2\cdot 8 + (2 + 0{,}725)^2\cdot 5) = (1/40)(64{,}354 + 62{,}108 +$$

$$+ 26{,}781 + 39{,}605 + 37{,}128) = 5{,}750 = 5{,}8\,;$$

$$D_B(Y) = \frac{1}{n}\sum_{j=1}^{l} n_j (y_j - \overline{y})^2 = (1/40)((-1 - 3{,}6)^2\cdot 10 + (2 - 3{,}6)^2\times$$

$$\times 14 + (7 - 3{,}6)^2\cdot 9 + (9 - 3{,}6)^2\cdot 7) = (1/40)(211{,}6 + 35{,}84 +$$

$$+ 104{,}04 + 204{,}12) = 13{,}89 = 13{,}9\,; \quad \sigma_B(X) = \sqrt{D_B(X)} = \sqrt{5{,}750} =$$

$$= 2{,}398 = 2{,}4\,; \quad \sigma_B(Y) = \sqrt{D_B(Y)} = \sqrt{13{,}89} = 3{,}727 = 3{,}7\,;$$

$$\overline{x}(y_j) = \frac{1}{n_j}\sum_{i} x_i\, q_{ij}\,; \quad \overline{x}(y_1) = (1/10)(1\cdot 2 + 1{,}5\cdot 5 + 2\cdot 3) = 1{,}55 =$$

$$= 1{,}6\,; \quad \overline{x}(y_2) = (1/14)(-3\cdot 3 + 1\cdot 6 + 1{,}5\cdot 3 + 2\cdot 2) = 0{,}393 = 0{,}4\,;$$

$$\overline{x}(y_3) = (1/9)(-3\cdot 8 + 1\cdot 1) = -2{,}556 = -2{,}6\,;$$

$$\overline{x}(y_4) = (1/7)(-4\cdot 6 + (-3)\cdot 1) = -3{,}857 = -3{,}9\,;$$

$$\overline{y}(x_i) = \frac{1}{m_i}\sum_{j} y_j\, q_{ij}\,; \quad \overline{y}(x_1) = (1/6)(9\cdot 6) = 9\,; \quad \overline{y}(x_2) = (1/12)\times$$

$$\times(2\cdot 3 + 7\cdot 8 + 9\cdot 1) = 5{,}917 = 5{,}9\,; \quad \overline{y}(x_3) = (1/9)(-1\cdot 2 + 2\cdot 6 +$$

$$+ 7\cdot 1) = 1{,}889 = 1{,}9\,; \quad \overline{y}(x_4) = (1/8)(-1\cdot 5 + 2\cdot 3) = 0{,}125 = 0{,}1\,;$$

$$\overline{y}(x_5) = (1/5)(-1\cdot 3 + 2\cdot 2) = 0{,}2\,; \quad \overline{xy} = (1/n)\sum_{i,j} x_i y_j\, q_{ij}\,;$$

$$\overline{xy}=(1/40)(-4\cdot 9\cdot 6+(-3)\cdot 2\cdot 3+(-3)\cdot 7\cdot 8+(-3)\cdot 9\cdot 1+$$
$$+1\cdot(-1)\cdot 2+1\cdot 2\cdot 6+1\cdot 7\cdot 1+1{,}5\cdot(-1)\cdot 5+$$
$$+1{,}5\cdot 2\cdot 3+2\cdot(-1)\cdot 3+2\cdot 2\cdot 2)=(1/40)(-216-18-168-$$
$$-27-2+12+7-7{,}5+9-6+8)=-10{,}213=-10{,}2\,;$$

$$r_B=\frac{\overline{xy}-\bar{x}\cdot\bar{y}}{\sigma_B(X)\cdot\sigma_B(Y)}=\frac{-10{,}213-(-0{,}725)\cdot 3{,}6}{2{,}398\cdot 3{,}727}=-0{,}851=-0{,}9\,;$$

$$A=r_B\cdot\sigma_B(Y)/\sigma_B(X)=-0{,}851\cdot 3{,}727/2{,}398=-1{,}323=-1{,}3\,;$$

$$B=\bar{y}-A\bar{x}=3{,}6-(-1{,}323)\cdot(-0{,}725)=2{,}641=2{,}6\,;$$

$$y=-1{,}323x+2{,}641\,;$$

x	-4	2
y	7,9	0,0

Складаємо таблицю координат вершин емпіричної ламаної регресії Y на X :

x_i	-4	-3	1	1,5	2
$\bar{y}(x_i)$	9	5,9	1,9	0,1	0,2

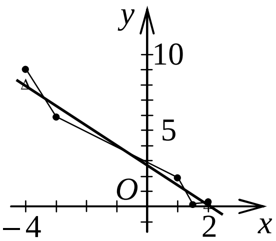


Рис. 38

На координатній площині Oxy будуємо емпіричну ламану регресії Y на X і емпіричну пряму регресії (рис. 38). ■

Зауваження 3. Вибірковий коефіцієнт кореляції r_B, що відображає тісноту лінійного зв’язку між вибірковими значеннями величин X і Y, за абсолютною величиною не перевищує одиниці: $|r_B|\le 1$. Якщо він дорівнює нулю, то ці значення X і Y не зв’язані лінійною кореляційною залежністю, при цьому вибіркова пряма регресії Y на X паралельна осі Ox. (Проте може спостерігатися нелінійна кореляційна чи функціональна залежність). При зростанні абсолютної величини r_B лінійна кореляційна залежність стає все більш тісною і при $|r_B|=1$ переходить у функціональну. Якщо вибірка репрезентативна (тобто, має достатній об’єм і добре відображає генеральну сукупність), то висновок про тісноту лінійного зв’язку між величинами X і Y, одержаний за даними вибірки, в певній мірі можна поширити і на генеральну сукупність.

2.6. Перевірка статистичних гіпотез

2.6.1. Статистична гіпотеза. Статистичний критерій. Помилки першого та другого роду. Критична область. Критичні точки

Будь-який висновок, отриманий на основі обробки статистичних даних, носить імовірнісний характер і служить лише обґрунтованим науковим припущенням, оскільки не є повністю достовірним.

Нехай вибірковим методом досліджується генеральна сукупність, пов’язана з деякою випадковою величиною X. Довільне висловлення (припущення) про генеральну сукупність (випадкову величину X), що перевіряють за вибіркою (тобто за результатами спостережень), називають ***статистичною гіпотезою*** H (або просто ***гіпотезою***). Гіпотези можуть бути ***параметричними*** – припущення про параметри відомого закону розподілу, і ***непараметричними*** – припущення про вигляд невідомого закону розподілу. Розрізняють ***прості*** гіпотези, що включають лише одне припущення, і ***складні***, що містять більше одного припущення. Основну гіпотезу, висунуту за результатами обробки статистичного матеріалу, називають ***нульовою*** і позначають H_0. На противагу їй звичайно розглядають одну або декілька ***альтернативних*** (***конкуруючих***) гіпотез, які позначають H_1, H_2, …. Процедуру зіставлення гіпотези з вибірковими даними (чи не суперечить висунута гіпотеза H_0 вибірці) називають ***перевіркою гіпотези***. Якщо висунута гіпотеза H_0 відкидається, то її місце займає одна з альтернативних.

Зауваження 1. Вибір альтернативних гіпотез визначається конкретним формулюванням задачі. Надалі обмежимося випадком лише однієї конкуруючої гіпотези H_1.

Довільне правило, за яким приймається чи відхиляється висунута гіпотеза H_0, називають ***статистичним критерієм*** (чи просто ***критерієм***) K перевірки гіпотези H_0. Розробка таких правил та їх обґрунтування з точки зору вимог оптимальності служить предметом теорії перевірки статистичних гіпотез.

Перевірка статистичних гіпотез ґрунтується на ***принципі практичної впевненості***, згідно з яким малоймовірні події вважають неможливими, а події з близькою до одиниці ймовірністю – достовірними.

Статистичними методами гіпотезу можна лише обґрунтовано відхилити чи прийняти, але не довести як безсумнівний факт. Тобто, результат перевірки гіпотези може виявитися помилковим. Прийнято розрізняти помилки наступних двох видів.

Помилка першого роду – відкинута нульова гіпотеза H_0, у той час як вона вірна ("пропуск цілі"). Ймовірність помилки першого роду позначають α і називають ***рівнем значущості***: $\alpha = P(H_1 / H_0)$. На практиці прийнятний рівень значущості α задають наперед (звичайно з діапазону $[0{,}01; 0{,}1]$).

Помилка другого роду – прийнята нульова гіпотеза H_0, у той час як вона невірна, тобто насправді справджується альтернативна гіпотеза H_1 ("хибне спрацювання"). Ймовірність такої помилки позначають β: $\beta = P(H_0 / H_1)$. Ймовірність $1-\beta$ не допустити помилку другого роду називають ***потужністю критерію***. Чим вона більша, тим кращий критерій, тим вище надійність перевірки.

Наступна таблиця наочно ілюструє можливі помилки.

Гіпотеза H_0	Відхиляється	Приймається
Вірна	Помилка 1 – го роду	Правильне рішення
Невірна	Правильне рішення	Помилка 2 – го роду

<u>Зауваження 2</u>. При вибраному критерії та заданому рівні значущості α ймовірність β помилки другого роду можна зменшити лише за рахунок збільшення об'єму вибірки.

Загальний підхід до побудови критеріїв полягає в наступному. Розглядають вибірку $x_1, x_2, ..., x_n$ об'єму n як спостережені значення сукупності незалежних випадкових величин $X_1, X_2, ..., X_n$, кожна з яких має такий же розподіл, що й випадкова величина X. Формують яку-небудь випадкову величину $U = U(X_1, X_2, ..., X_n)$ – ***статистику критерію***, що характеризує відхилення емпіричних даних $x_1, x_2, ..., x_n$ від відповідних (згідно гіпотези H_0) гіпотетичних значень і розподіл якої у випадку справедливості H_0 можна визначити.

Виходячи з припущення, що дійсно вірна гіпотеза H_0, задають прийнятний рівень значущості α і розбивають всю область значень статистики критерію U на дві підмножини S_α и $S_{1-\alpha}$, що не мають спільних елементів. Першу з них S_α називають ***критичною областю***, при потраплянні в яку гіпотезу H_0 відхиляють, а другу $S_{1-\alpha}$ – ***областю прийняття гіпотези*** H_0. Значення статистики критерію, що відокремлюють ці області одну від одної, називають ***критичними точками***. Якщо фактичне значення u статистики критерію U, обчислене за вибіркою, потрапляє в критичну область S_α: $u \in S_\alpha$, то основну гіпотезу H_0 відхиляють і приймають альтернативну гіпотезу H_1. Якщо ж $u \in S_{1-\alpha}$, то навпаки, основну гіпотезу H_0 приймають і відкидають конкуруючу гіпотезу H_1.

Можливі три випадки розміщення критичної області S_α, в залежності від вигляду нульової та конкуруючої гіпотез і закону розподілу статистики критерію U.

Правостороння критична область (рис. 39) – це інтервал $(u_{кр2,\alpha};+\infty)$, де $u_{кр2,\alpha}$ – ***правостороння критична точка***, що відповідає рівню значущості α і визначається з умови

$$P(U > u_{кр2,\alpha}) = \alpha.$$

Лівостороння критична область (рис. 40) – це інтервал $(-\infty;u_{кр1,\alpha})$, де $u_{кр1,\alpha}$ – ***лівостороння критична точка***, що відповідає рівню значущості α і визначається з умови

$$P(U < u_{кр1,\alpha}) = \alpha.$$

Двостороння критична область (рис. 41) складається з двох інтервалів $(-\infty;u_{кр1,\alpha/2})$ і $(u_{кр2,\alpha/2};+\infty)$, де $u_{кр1,\alpha/2}$ і $u_{кр2,\alpha/2}$ – ***ліва та права критичні точки***, що відповідають рівню значущості α і визначаються з умов

$$P(U < u_{кр1,\alpha/2}) = \alpha/2 \quad \text{і} \quad P(U > u_{кр2,\alpha/2}) = \alpha/2 .$$

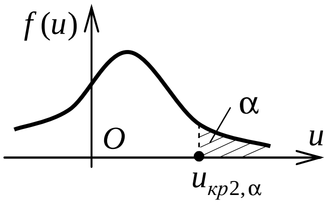


Рис. 39

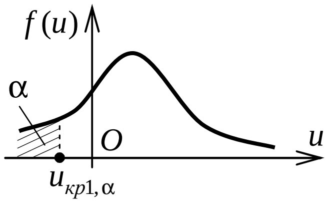


Рис. 40

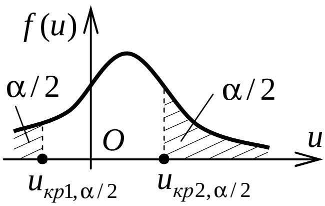


Рис. 41

Зауваження 3. Статистику критерію $U = U(X_1, X_2, ..., X_n)$ звичайно підбирають так, щоб ця випадкова величина U мала один з наступних стандартних розподілів: стандартний нормальний розподіл, розподіл Пірсона (χ^2-розподіл), розподіл Стьюдента (t-розподіл), розподіл Фішера – Снедекора (F-розподіл). Який з них варто використовувати, залежить від характеру задачі.

Для практичного застосування критерій перевірки гіпотез можна конкретизувати і подати у вигляді наступної схеми:

1) Сформулювати основну H_0 й альтернативну H_1 гіпотези.

2) Задати прийнятний рівень значущості α.

3) Вибрати статистику критерію $U = U(X_1, X_2, ..., X_n)$ для перевірки нульової гіпотези H_0, керуючись специфікою генеральної сукупності (випадкової величини X).

4) Визначити вибірковий розподіл статистики критерію $U = U(X_1, X_2, ..., X_n)$ при умові, що вірна гіпотеза H_0.

5) Визначити відповідну критичну область S_α. Для цього досить знайти критичні точки, що визначають її межу. Для кожної стандартної статистики існують довідкові таблиці та комп'ютерні процедури, за якими знаходять ці точки.

6) Одержати вибірку $x_1, x_2, ..., x_n$ спостережень і обчислити за нею

фактичне значення u статистики критерію U.

7) Прийняти статистичне рішення: якщо $u \in S_\alpha$, то гіпотеза H_0 відхиляється, оскільки вона не узгоджується з результатами спостережень; якщо $u \notin S_\alpha$, то гіпотеза H_0 приймається, оскільки вона не суперечить результатам спостережень.

Далі розглянемо ряд типових задач перевірки гіпотез.

2.6.2. Перевірка гіпотези про математичне сподівання нормально розподіленої випадкової величини

Нехай випадкова величина X, що визначена на множині об'єктів деякої генеральної сукупності, розподілена нормально з відомою дисперсією $D(X) = \sigma^2$, а її математичне сподівання $M(X)$ невідоме.

Припустимо, що є деякі підстави вважати, що $M(X) = a$, де a – деяке число. Такими підставами можуть служити: накопичений досвід дослідження подібних випадкових величин, певні відомості про об'єкти генеральної сукупності, зокрема, одержані за результатами вибіркових досліджень і т.п. Тоді висуваємо нульову гіпотезу $\boxed{H_0:\ M(X) = a}$ при конкуруючій гіпотезі $\boxed{H_1:\ M(X) \neq a}$.

Здійснюємо вибірку $x_1, x_2, ..., x_n$ об'єму n. Враховуючи, що вибіркове середнє $\bar{x}$ як випадкова величина розподілене за нормальним законом з дисперсією $D(\bar{x}) = D(X)/n = \sigma^2/n$ і математичним сподіванням $M(\bar{x}) = M(X)$, можна записати нульову гіпотезу H_0 так: $M(\bar{x}) = a$. Для її перевірки можна вибрати статистику $\boxed{Z = (\bar{x} - a)/\sigma(\bar{x}) = (\bar{x} - a)\sqrt{n}/\sigma}$ як міру розбіжності між вибірковим $\bar{x}$ і генеральним a середніми. Ця випадкова величина також має нормальний розподіл, причому, якщо нульова гіпотеза H_0 вірна, то $M(Z) = 0$ і $\sigma(Z) = 1$. Подамо відповідну альтернативну гіпотезу H_1: $M(\bar{x}) \neq a$. При такому формулюванні конкуруючої гіпотези значні відхилення величини Z в обидва боки від нуля повинні приводити до висновку про хибність нульової гіпотези H_0.

Тому необхідно побудувати двосторонню критичну область S_α, що відповідає прийнятому рівню значущості α. При цьому найбільша потужність критерію досягається, якщо ймовірність влучення статистики критерію Z у кожний з двох інтервалів області S_α дорівнює $\alpha/2$. Оскільки розподіл Z симетричний відносно нуля, критичні точки також розташовані симетрично відносно нуля, тобто $(-z_{кр,\alpha/2}; z_{кр,\alpha/2})$ – область прийняття гіпотези H_0, де $z_{кр,\alpha/2}$ – критична точка, що визначається з умови $P(|Z| > z_{кр,\alpha/2}) = \alpha/2$.

Користуючись функцією Лапласа $\Phi(x)$, критичну точку $z_{кр,\alpha/2}$ можна знайти як корінь рівняння $\Phi(z_{кр,\alpha/2}) = (1-\alpha)/2$.

Нехай $z = (\bar{x} - a)\sqrt{n}\big/\sigma$ – фактичне значення статистики критерію Z, одержане за вибіркою. Тоді при $|z| < z_{кр,\alpha/2}$ нульова гіпотеза H_0 приймається, а при $|z| > z_{кр,\alpha/2}$ ця гіпотеза відхиляється.

Коли <u>генеральна дисперсія</u> $D(X)$ <u>невідома</u>, за статистику критерію приймають випадкову величину $\boxed{T = (\bar{x} - a)\sqrt{n}\big/S}$, де S – виправлене вибіркове середнє квадратичне відхилення. Випадкова величина T має t-розподіл (розподіл Стьюдента) з $k = n-1$ ступенями свободи. При цьому також розглядають симетричну двосторонню критичну область $S_\alpha : |t| > t_{двост,\alpha,k}$, де $t_{двост,\alpha,k}$ – критична точка розподілу Стьюдента при двосторонній критичній області, що відповідає прийнятому рівню значущості α.

Нехай $t = (\bar{x} - a)\sqrt{n}\big/S$ – емпіричне значення статистики критерію T, обчислене за вибіркою. Тоді при $|t| < t_{двост,\alpha,k}$ нульова гіпотеза H_0 приймається, а при $|t| > t_{двост,\alpha,k}$ ця гіпотеза відкидається.

<u>Приклад</u>. Нехай випадкова величина X розподілена за нормальним законом з невідомими дисперсією $D(X)$ і математичним сподіванням $M(X)$. Попередні дослідження дають підстави вважати, що $M(X) = a = 6$. За вибіркою $x_1, x_2, ..., x_n$ об'єму $n = 49$

знайдені вибіркове середнє $\overline{x} = 5{,}8$ і виправлене вибіркове середнє квадратичне відхилення $S = 0{,}8$. Необхідно при рівні значущості $\alpha = 0{,}05$ перевірити нульову гіпотезу $H_0 :\ M(X) = 6$ при альтернативній гіпотезі $H_1 :\ M(X) \neq 6$.

□ Обчислимо фактичне значення статистики критерію T:

$$t = (\overline{x} - a)\sqrt{n}\big/S = (5{,}8 - 6)\sqrt{49}\big/0{,}8 = -1{,}75.$$

За таблицею розподілу Стьюдента з $k = n - 1 = 49 - 1 = 48$ ступенями свободи при $\alpha = 0{,}05$ знайдемо двосторонню критичну точку: $t_{двост,\alpha,k} = t_{двост,0{,}05;48} = 2{,}01$. Оскільки

$$|t| = 1{,}75 < t_{двост,0{,}05;48} = 2{,}01,$$

то нульова гіпотеза $H_0 :\ M(X) = 6$ приймається. Тобто, можна вважати, що відхилення вибіркового середнього $\overline{x} = 5{,}8$ від очікуваного значення $a = 6$ математичного сподівання $M(X)$ незакономірне і зумовлене випадковими факторами. ■

2.6.3. Перевірка гіпотези про рівність математичних сподівань двох нормально розподілених випадкових величин

У практиці статистичних досліджень часто виникають ситуації, коли середній результат для однієї серії випробувань відрізняється від аналогічного для іншої. Це породжує питання, чи можна виявлену розбіжність пояснити неминучими випадковими похибками експерименту, чи вона обумовлена деякими невиявленими закономірностями. Зокрема, подібні випадки спостерігаються при проведенні вибіркового контролю якості виробів, що виготовляються або різними працівниками, або на різному обладнанні чи при різних технологічних режимах.

Нехай випадкові величини X і Y розподілені нормально з відомими дисперсіями $D(X)$ і $D(Y)$, проте з невідомими математичними сподіваннями $M(X)$ і $M(Y)$. З обох генеральних сукупностей X і Y витягнуті незалежні вибірки об’єму відповідно m і n, за якими обчислені вибіркові середні $\overline{x}$ і $\overline{y}$. Необхідно за емпіричними середніми $\bar{x}$ і $\bar{y}$ при заданому рівні значущості α перевірити нульову гіпотезу $\boxed{H_0:\ M(X)=M(Y)}$ про рівність математичних сподівань при конкуруючій гіпотезі $\boxed{H_1:\ M(X)\neq M(Y)}$.

Статистикою критерію служить нормально розподілена випадкова величина $\boxed{Z=\dfrac{\bar{x}-\bar{y}}{\sqrt{(1/m)D(X)+(1/n)D(Y)}}}$, причому, якщо нульова гіпотеза H_0 вірна, то $M(Z)=0$ і $\sigma(Z)=1$. Критична область S_α є двосторонньою симетричною і задається нерівністю $|Z|>z_{кр,\alpha/2}$, де $z_{кр,\alpha/2}$ – критична точка, що визначається як корінь рівняння $\Phi(z_{кр,\alpha/2})=(1-\alpha)/2$.

Нехай $z=\dfrac{\bar{x}-\bar{y}}{\sqrt{(1/m)D(X)+(1/n)D(Y)}}$ – фактичне значення статистики критерію Z, одержане за вибірками з X і Y. Тоді при $|z|<z_{кр,\alpha/2}$ нульова гіпотеза H_0 приймається, а при $|z|>z_{кр,\alpha/2}$ ця гіпотеза відхиляється.

Зауваження 1. Якщо вибірки $x_1,x_2,...,x_m$ і $y_1,y_2,...,y_n$ незалежні й обидві мають значний об'єм ($m>30$ і $n>30$), то наведене правило можна застосовувати навіть тоді, коли генеральні сукупності X і Y не мають нормального розподілу та їх дисперсії $D(X)$ і $D(Y)$ невідомі. При цьому можна вважати, що вибіркові середні $\bar{x}$ і $\bar{y}$ як випадкові величини розподілені приблизно нормально, а вибіркові дисперсії $D_B(X)$ і $D_B(Y)$ служать досить добрими оцінками генеральних дисперсій $D(X)$ і $D(Y)$. За статистику критерію Z можна наближено прийняти $\boxed{Z=\dfrac{\bar{x}-\bar{y}}{\sqrt{(1/m)D_B(X)+(1/n)D_B(Y)}}}$.

Зауваження 2. Нехай випадкові величини X і Y розподілені нормально, причому їх дисперсії $D(X)$ і $D(Y)$ невідомі, а незалежні вибірки $x_1,x_2,...,x_m$ і $y_1,y_2,...,y_n$ мають малий об'єм, що не дозволяє одержати хороші вибіркові оцінки $D_B(X)$ і $D_B(Y)$ генеральних дисперсій $D(X)$ і $D(Y)$. Якщо припустити рівність між собою невідомих генеральних дисперсій $D(X)=D(Y)$, то за статистику критерію можна прийняти випадкову величину

$$\boxed{T=\frac{\bar{x}-\bar{y}}{\sqrt{(m-1)S^2(X)+(n-1)S^2(Y)}}\sqrt{\frac{mn(m+n-2)}{m+n}}},$$

яка при справедливості нульової гіпотези H_0 має розподіл Стьюдента з $k=m+n-2$ ступенями свободи. Тут $S^2(X)$ і $S^2(Y)$ – виправлені вибіркові дисперсії. При цьому розглядають симетричну двосторонню критичну область $S_\alpha : |t| > t_{двост,\alpha,k}$, де $t_{двост,\alpha,k}$ – критична точка розподілу Стьюдента, що відповідає прийнятому рівню значущості α. Якщо емпіричне значення t статистики критерію T, обчислене за вказаними вибірками, задовольняє нерівність $|t| < t_{двост,\alpha,k}$, то нульова гіпотеза H_0 приймається, а у випадку $|t| > t_{двост,\alpha,k}$ ця гіпотеза відхиляється і приймається конкуруюча гіпотеза H_1.

Зауваження 3. Останній критерій слабо чутливий до припущення про нормальність розподілів випадкових величин X і Y. Його можна застосовувати, коли ці розподіли є одномодальними і не дуже асиметричними. Припущення $D(X)=D(Y)$ у багатьох випадках може бути обґрунтоване на предметному рівні (наприклад, дисперсії обох генеральних сукупностей визначаються похибками вимірювального приладу). Коли ж немає вагомих підстав вважати дисперсії однаковими, то треба перевірити гіпотезу $D(X)=D(Y)$, застосовуючи відповідний критерій (див. п. 2.6.4).

Приклад. Нехай випадкові величини X і Y розподілені нормально і мають однакові, хоча й невідомі, дисперсії $D(X)=D(Y)$. За незалежними вибірками $x_1, x_2, ..., x_m$ і $y_1, y_2, ..., y_n$ об'єму відповідно $m=25$ і $n=15$ знайдені вибіркові середні $\bar{x}=3{,}2$ і $\bar{y}=2{,}7$ та виправлені вибіркові дисперсії $S^2(X)=2{,}1$ і $S^2(Y)=1{,}9$, Необхідно при рівні значущості $\alpha=0{,}01$ перевірити нульову гіпотезу

$H_0: M(X) = M(Y)$ про рівність математичних сподівань при альтернативній гіпотезі $H_1: M(X) \neq M(Y)$.

□ Обчислимо фактичне значення t статистики критерію T:

$$t = \frac{3{,}2 - 2{,}7}{\sqrt{(25-1)\cdot 2{,}1 + (15-1)\cdot 1{,}9}} \sqrt{\frac{25\cdot 15\cdot (25+15-2)}{25+15}} = 1{,}08.$$

За таблицею розподілу Стьюдента з $k = m + n - 2 = 25 + 15 - - 2 = 38$ ступенями свободи при $\alpha = 0{,}01$ знайдемо двосторонню критичну точку: $t_{двост,\alpha,k} = t_{двост,0{,}01;38} = 2{,}70$. Оскільки

$$|t| = 1{,}08 < t_{двост,0{,}01;38} = 2{,}70,$$

то нульова гіпотеза $H_0: M(X) = M(Y)$ приймається. Тобто, при рівні значущості $\alpha = 0{,}01$ можна вважати, що математичні сподівання випадкових величин X і Y рівні між собою. ■

2.6.4. Перевірка гіпотези про рівність дисперсій двох нормально розподілених випадкових величин

Гіпотези про дисперсію відіграють дуже важливу роль, оскільки величина розсіювання, що характеризується дисперсією, емпіричних вибіркових даних відносно обчислених значень відповідних параметрів дозволяє судити про придатність (адекватність) теоретичних положень, на яких ґрунтуються ці розрахунки.

На практиці необхідність порівняння дисперсій виникає при дослідженні якості налагодження обладнання, стійкості технологічних процесів, точності приладів та інструментів і т.п.

Нехай з нормально розподілених генеральних сукупностей X і Y здійснені незалежні вибірки $x_1, x_2, ..., x_m$ і $y_1, y_2, ..., y_n$ об'єму відповідно m і n, за якими обчислені виправлені вибіркові дисперсії $S^2(X)$ і $S^2(Y)$, що звичайно є різними. Припустимо, для визначеності, $\boxed{S^2(X) \geq S^2(Y)}$. Нехай варто встановити, чи взяті розглянуті вибірки з нормальних генеральних сукупностей X і Y з однаковою дисперсією $D(X) = D(Y)$.

Необхідно при заданому рівні значущості α за емпіричними оцінками $S^2(X)$ і $S^2(Y)$ генеральних дисперсій $D(X)$ і $D(Y)$ перевірити нульову гіпотезу $\boxed{H_0:\ D(X)=D(Y)}$ про рівність останніх при альтернативній гіпотезі $\boxed{H_1:\ D(X)>D(Y)}$, яка полягає в тому, що генеральна дисперсія більша у тієї випадкової величини X, в якої більша виправлена вибіркова дисперсія.

Статистикою критерію служить випадкова величина F, що є відношенням більшої виправленої вибіркової дисперсії $S^2(X)$ до меншої $S^2(Y)$: $\boxed{F=S^2(X)\big/S^2(Y)}$, яка при справедливості нульової гіпотези H_0 має F-розподіл (розподіл Фішера – Снедекора) з $k_1=m-1$ і $k_2=n-1$ ступенями свободи.

У цій задачі природно розглядати правосторонню критичну область S_α, що задається нерівністю $F>F_{кр,\alpha,k_1,k_2}$, де $F_{кр,\alpha,k_1,k_2}$ – правостороння критична точка розподілу Фішера – Снедекора, що відповідає прийнятому рівню значущості α.

Якщо емпіричне значення $f=S^2(X)\big/S^2(Y)$ статистики критерію F, обчислене за вказаними вибірками, задовольняє нерівність $f<F_{кр,\alpha,k_1,k_2}$, то нульова гіпотеза H_0 приймається, а у випадку $f>F_{кр,\alpha,k_1,k_2}$ ця гіпотеза відкидається і приймається конкуруюча гіпотеза H_1.

Приклад. Нехай відносно певної спільної кількісної ознаки проводиться вибірковий контроль двох множин – генеральних сукупностей X і Y – деяких деталей, що виготовлені відповідно на першому та другому верстатах. Відомо, що генеральні сукупності X і Y розподілені за нормальним законом. За результатами досліджень сформовані дві незалежні вибірки $x_1,x_2,...,x_m$ і $y_1,y_2,...,y_n$ об'єму відповідно $m=9$ і $n=11$, за якими обчислені виправлені вибіркові дисперсії $S^2(X)=1{,}55$ і $S^2(Y)=0{,}45$. Необхідно при рівні значущості $\alpha=0{,}05$ перевірити нульову гіпотезу H_0 про рівність генеральних дисперсій $D(X)=D(Y)$ при конкуруючій гіпотезі H_1: $D(X) > D(Y)$.

□ Обчислимо фактичне значення f статистики критерію F:

$$f = S^2(X) \big/ S^2(Y) = 1{,}55/0{,}45 = 3{,}44 .$$

За таблицею F-розподілу з $k_1 = m - 1 = 9 - 1 = 8$ і $k_2 = n - 1 = 11 - 1 = 10$ ступенями свободи при $\alpha = 0{,}05$ знайдемо критичну точку: $F_{кр,\alpha,k_1,k_2} = F_{кр;0{,}05;8;10} = 3{,}07$. Оскільки

$$f = 3{,}44 > F_{кр;0{,}05;8;10} = 3{,}07 ,$$

то нульова гіпотеза H_0 про рівність генеральних дисперсій $D(X)$ і $D(Y)$ відхиляється. Тобто, при рівні значущості $\alpha = 0{,}05$ можна вважати, що справджується альтернативна гіпотеза H_1: дисперсія $D(X)$ випадкової величини X значущо перевищує дисперсію $D(Y)$ випадкової величини Y. ■

2.6.5. Перевірка гіпотези про нормальний закон розподілу випадкової величини

У багатьох практичних задачах при дослідженні випадкової величини X навіть вигляд її закону розподілу наперед невідомий. Якщо закон розподілу генеральної сукупності визначається за вибіркою, то виникає необхідність оцінити, якою є розбіжність між емпіричним і прийнятим теоретичним розподілами – випадковою чи значущою. Критерій перевірки гіпотези про гаданий закон невідомого розподілу називають ***критерієм згоди***.

Найпоширеніший критерій згоди – це ***критерій*** χ^2 (***критерій Пірсона***), який може використовуватися для перевірки гіпотези про довільний закон розподілу. Розглянемо його застосовування для перевірки гіпотези про нормальний закон розподілу випадкової величини X.

Нехай з генеральної сукупності X з невідомою функцією розподілу $F(x)$ здійснено вибірку $x_1, x_2, ..., x_n$ досить великого об'єму n ($n \ge 50$) зі значним числом різних варіант, за якою сформовано інтервальний статистичний ряд

i	1	2	$\ldots$	$m-1$	m	–
x_i	$[a_0;a_1)$	$[a_1;a_2)$	$\ldots$	$[a_{m-2};a_{m-1})$	$[a_{m-1};a_m]$	$\sum_{i=1}^{m}$
n_i	n_1	n_2	$\ldots$	n_{m-1}	n_m	n

так, що $n_i \geq 5$, $i=\overline{1,m}$, де m – число елементарних інтервалів (у загальному випадку, нерівномірних).

Виходячи з рівня значущості α, висувається нульова гіпотеза $\boxed{H_0: X \text{ має нормальний закон розподілу}}$ при альтернативній гіпотезі $\boxed{H_1: X \text{ має відмінний від нормального закон розподілу}}$.

Зауваження. Якщо виявиться, що для деякого i-го частинного проміжку умова $n_i \geq 5$ не виконується, то його треба об'єднати з сусіднім інтервалом.

Статистикою критерію служить випадкова величина $\boxed{\chi^2 = \sum_{i=1}^{m}(n_i - n_{T\,i})^2 \big/ n_{T\,i}}$, де n_i і $n_{T\,i}$ – відповідно емпірична (спостережена) і теоретична частота попадання вибіркових значень випадкової величини X в i-й частинний інтервал з кінцями a_{i-1} і a_i, $i=\overline{1,m}$. Теоретична частота $n_{T\,i}$ обчислюється як кількість вибіркових значень, які повинні потрапити в i-й інтервал при умові, що випадкова величина X розподілена за прийнятим нормальним законом, параметри якого співпадають з їх точковими оцінками за цією ж вибіркою: $\boxed{a=\bar{x}}$ і $\boxed{\sigma=\sigma_B}$. Нехай

$$\boxed{p_i = P(a_{i-1} < X < a_i) = \Phi\big((a_i-\bar{x})/\sigma_B\big) - \Phi\big((a_{i-1}-\bar{x})/\sigma_B\big)}$$

– теоретична ймовірність попадання випадкової величини X, що за припущенням має вибраний нормальний розподіл, в i-й частинний інтервал. Тоді дістанемо $\boxed{n_{T\,i} = n\,p_i}$. Тут $\Phi(x)$ – функція Лапласа.

Статистика критерію χ^2 характеризує близькість емпіричного й теоретичного розподілів: чим краще узгоджені ці розподіли, тим менше розрізняються емпіричні й теоретичні частоти і, відповідно, менше значення статистики критерію.

Необхідно знати розподіл статистики критерію χ^2 як випадкової величини у припущенні, що нульова гіпотеза H_0 справджується. Точний розподіл випадкової величини χ^2 незручний для обчислень. Проте у випадку вибірки великого об'єму n ($n \geq 50$), для якої $n_i \geq 5$ ($i = \overline{1, m}$), статистика критерію χ^2 при вірності гіпотези H_0 наближено має досить простий χ^2-розподіл. Пірсон довів, що *незалежно від того, який розподіл реально має випадкова величина* X, *закон розподілу статистики критерію* χ^2 *при* $n \to \infty$ *прямує до* χ^2*- розподілу з* $k = m - 1 - s$ *ступенями свободи, де* s *– число параметрів прийнятого за припущенням теоретичного закону розподілу, що оцінюються за даними вибірки.*

Оскільки нормальний закон характеризується двома параметрами a і σ, то для нього $\boxed{k = m - 3}$.

Для вибраної статистики критерію χ^2 розглядається правостороння критична область S_α, що задається нерівністю $\chi^2 > \chi^2_{кр,\alpha,k}$, де $\chi^2_{кр,\alpha,k}$ – правостороння критична точка χ^2- розподілу, що відповідає прийнятому рівню значущості α.

Якщо емпіричне значення $\chi^2_B = \sum_{i=1}^{m} (n_i - n_{T\,i})^2 \big/ n_{T\,i}$ статистики критерію χ^2, обчислене за вибіркою, задовольняє нерівність $\chi^2_B < \chi^2_{кр,\alpha,k}$, то нульова гіпотеза H_0 приймається, тобто за даними спостережень випадкова величина X має нормальний закон розподілу, розбіжність між емпіричними і теоретичними частотами носить випадковий характер і є незначущою. У випадку $\chi^2_B > \chi^2_{кр,\alpha,k}$ гіпотеза H_0 відхиляється і приймається конкуруюча гіпотеза H_1, тобто випадкова величина X має відмінний від нормального закон розподілу, розбіжність між емпіричними і теоретичними частотами суттєва.

Приклад. Для вибірки $x_1, x_2, ..., x_n$ з генеральної сукупності X, інтервальний статистичний ряд якої подано наступною таблицею, необхідно при рівні значущості $\alpha = 0{,}05$ перевірити нульову гіпотезу H_0: X має нормальний закон розподілу при альтернативній гіпотезі H_1: X має відмінний від нормального закон розподілу за допомогою критерію Пірсона.

Номер інтервалу i	Ліва межа інтервалу a_{i-1}	Права межа інтервалу a_i	Емпірична частота n_i
1	-3	1	5
2	1	5	6
3	5	8	9
4	8	10	18
5	10	12	22
6	12	15	19
7	15	17	14
8	17	20	5
9	20	22	2

□ Оскільки $n_9 = 2 < 5$, то дев'ятий інтервал об'єднаємо з сусіднім. Тоді статистичний ряд набуде вигляду:

Номер інтервалу i	Ліва межа інтервалу a_{i-1}	Права межа інтервалу a_i	Емпірична частота n_i
1	-3	1	5
2	1	5	6
3	5	8	9
4	8	10	18
5	10	12	22
6	12	15	19
7	15	17	14
8	17	22	7

Число частинних інтервалів $m = 8$, число ступенів свободи $k = m - 3 = 8 - 3 = 5$, об'єм вибірки

$$n = \sum_{i=1}^{m} n_i = 5 + 6 + 9 + 18 + 22 + 19 + 14 + 7 = 100.$$

Будемо вважати варіантами $x_1, x_2, ..., x_m$ середини елементарних інтервалів $x_i = (a_{i-1} + a_i)/2$:

$$x_1=(-3+1)/2=-1;\;\; x_2=(1+5)/2=3;\;\; x_3=(5+8)/2=6{,}5;$$

$$x_4=(8+10)/2=9;\; x_5=(10+12)/2=11;\; x_6=(12+15)/2=13{,}5;$$

$$x_7=(15+17)/2=16; \qquad x_8=(17+22)/2=19{,}5.$$

За одержаними варіантами знайдемо вибіркові оцінки параметрів гаданого нормального закону розподілу $a=\bar{x}$ і $\sigma=\sigma_B$:

$$\bar{x}=(1/n)\sum\nolimits_{i=1}^{m} n_i\,x_i=(5\cdot(-1)+6\cdot 3+9\cdot 6{,}5+18\cdot 9+22\cdot 11+19\times$$

$$\times 13{,}5+14\cdot 16+7\cdot 19{,}5)/100=10{,}9;\;\; a=\bar{x}=10{,}9;$$

$$D_B=(1/n)\sum\nolimits_{i=1}^{m} n_i\,(x_i-\bar{x})^2=(5\cdot(-1-10{,}9)^2+6\cdot(3-10{,}9)^2+$$

$$+9\cdot(6{,}5-10{,}9)^2+18\cdot(9-10{,}9)^2+22\cdot(11-10{,}9)^2+19\cdot(13{,}5-$$

$$-10{,}9)^2+14\cdot(16-10{,}9)^2+7\cdot(19{,}5-10{,}9)^2)/100=24{,}19;$$

$$\sigma_B=\sqrt{D_B}=\sqrt{24{,}19}=4{,}92;\quad \sigma=\sigma_B=4{,}92.$$

Обчислимо теоретичні ймовірності $n_{T\,i}$ і відповідні теоретичні частоти $n_{T\,i}=n\,p_i$ попадання випадкової величини X, в i-й частинний інтервал ($i=\overline{1,m}$) у припущенні про нормальний закон розподілу зі знайденими параметрами $a=\bar{x}=10{,}9$ і $\sigma=\sigma_B=4{,}92$:

$$p_i=\Phi\big((a_i-\bar{x})/\sigma_B\big)-\Phi\big((a_{i-1}-\bar{x})/\sigma_B\big);$$

$$p_1=\Phi\big((1-10{,}9)/4{,}92\big)-\Phi\big((-3-10{,}9)/4{,}92\big)=\Phi(-2{,}01)-$$

$$-\Phi(-2{,}83)=-\Phi(2{,}01)+\Phi(2{,}83)=-0{,}47778+0{,}49683=0{,}0191;$$

$$n_{T\,1}=100\cdot 0{,}0191=1{,}91;\qquad p_2=\Phi\big((5-10{,}9)/4{,}92\big)-\Phi\big((1-$$

$$-10{,}9)/4{,}92\big)=\Phi(-1{,}20)--\Phi(-2{,}01)=-\Phi(1{,}20)+\Phi(2{,}01)=$$

$$=-0{,}38493+0{,}47778=0{,}0929;\qquad n_{T\,2}=100\cdot 0{,}0929=9{,}29;$$

$$p_3=\Phi\big((8-10{,}9)/4{,}92\big)-\Phi\big((5-10{,}9)/4{,}92\big)=\Phi(-0{,}59)-$$

$$-\Phi(-1{,}20)=-\Phi(0{,}59)+\Phi(1{,}20)=-0{,}22240+0{,}38493=0{,}1625;$$

$$n_{T\,3}=100\cdot 0{,}1625=16{,}25;\qquad p_4=\Phi\big((10-10{,}9)/4{,}92\big)-\Phi\big((8-$$

$$-10{,}9)/4{,}92\big) = \Phi(-0{,}18) - \Phi(-0{,}59) = -\Phi(0{,}18) + \Phi(0{,}59) =$$

$$= -0{,}07142 + 0{,}22240 = 0{,}1510; \quad n_{T\,4} = 100 \cdot 0{,}1510 = 15{,}10;$$

$$p_5 = \Phi\big((12 - 10{,}9)/4{,}92\big) - \Phi\big((10 - 10{,}9)/4{,}92\big) = \Phi(0{,}22) -$$

$$-\Phi(-0{,}18) = \Phi(0{,}22) + \Phi(0{,}18) = 0{,}08706 + 0{,}07142 = 0{,}1585;$$

$$n_{T\,5} = 100 \cdot 0{,}1585 = 15{,}85; \quad p_6 = \Phi\big((15 - 10{,}9)/4{,}92\big) - \Phi\big((12 -$$

$$-10{,}9)/4{,}92\big) = \Phi(0{,}83) - \Phi(0{,}22) = 0{,}29673 - 0{,}08706 = 0{,}2097;$$

$$n_{T\,6} = 100 \cdot 0{,}2097 = 20{,}97; \quad p_7 = \Phi\big((17 - 10{,}9)/4{,}92\big) - \Phi\big((15 -$$

$$-10{,}9)/4{,}92\big) = \Phi(1{,}24) - \Phi(0{,}83) = 0{,}39251 - 0{,}29673 = 0{,}0958;$$

$$n_{T\,7} = 100 \cdot 0{,}0958 = 9{,}58; \quad p_8 = \Phi\big((22 - 10{,}9)/4{,}92\big) - \Phi\big((17 -$$

$$-10{,}9)/4{,}92\big) = \Phi(2{,}26) - \Phi(0{,}83) = 0{,}48809 - 0{,}39251 = 0{,}0956;$$

$$n_{T\,8} = 100 \cdot 0{,}0956 = 9{,}56.$$

Обчислимо за вибіркою фактичне значення χ^2_B статистики критерію χ^2:

$$\chi^2_B = \sum\nolimits_{i=1}^{m} (n_i - n_{T\,i})^2 \Big/ n_{T\,i} = (5 - 1{,}91)^2/1{,}91 + (6 - 9{,}29)^2/9{,}29 +$$

$$+ (9 - 16{,}25)^2/16{,}25 + (18 - 15{,}10)^2/15{,}10 + (22 - 15{,}85)^2/15{,}85 +$$

$$+ (19 - 20{,}97)^2/20{,}97 + (14 - 9{,}58)^2/9{,}58 + (7 - 9{,}56)^2/9{,}56 =$$

$$+ (19 - 20{,}97)^2/20{,}97 + (14 - 9{,}58)^2/9{,}58 + (7 - 9{,}56)^2/9{,}56 =$$

$$= 4{,}999 + 1{,}165 + 3{,}235 + 0{,}557 + 2{,}386 + 0{,}185 + 2{,}039 + 0{,}686 =$$

$$= 15{,}25.$$

За таблицею χ^2-розподілу з $k = 5$ ступенями свободи при $\alpha = 0{,}05$ знайдемо критичну точку: $\chi^2_{кр,\alpha,k} = \chi^2_{кр;0{,}05;5} = 11{,}1$.

Оскільки $\chi^2_B = 15{,}25 > \chi^2_{кр;0{,}05;5} = 11{,}1$, то гіпотеза H_0 про нормальний розподіл генеральної сукупності X відхиляється, розбіжність між емпіричними і теоретичними частотами значуща. ■

2.6.6. Перевірка гіпотези про значущість коефіцієнта кореляції

Нехай з нормально розподіленої двовимірної генеральної сукупності (X,Y) здійснено вибірку $(x_1,y_1),(x_2,y_2),...,(x_n,y_n)$ об'єму n, за якою обчислено вибірковий коефіцієнт кореляції r_B, що служить статистичною оцінкою коефіцієнта кореляції r генеральної сукупності. Якщо емпіричне значення r_B виявилося відмінним від нуля, то це ще не означає відмінність від нуля генерального коефіцієнта кореляції r. Тому перед тим, як робити висновок про наявність і ступінь корельованості випадкових величин X і Y, треба за вибіркою при заданому рівні значущості α перевірити нульову гіпотезу $\boxed{H_0:\ r=0}$ при конкуруючій гіпотезі $\boxed{H_1:\ r\neq 0}$.

Статистикою критерію служить випадкова величина

$$\boxed{T=r_B\sqrt{(n-2)\big/(1-r_B^2)}},$$

яка при справедливості нульової гіпотези H_0 має розподіл Стьюдента з $k=n-2$ ступенями свободи.

Оскільки r_B приймає значення з проміжку $[-1;1]$, то відносно великим відхиленням в обидва боки від нуля значенням вибіркового коефіцієнта кореляції r_B відповідають близькі до одиниці значення його модуля $|r_B|$. Тому природно розглядати симетричну двосторонню критичну область S_α, що задається нерівністю $|t|>t_{двост,\alpha,k}$, де $t_{двост,\alpha,k}$ – критична точка розподілу Стьюдента, що відповідає прийнятому рівню значущості α.

Якщо емпіричне значення $t=r_B\sqrt{(n-2)\big/(1-r_B^2)}$ статистики критерію T, обчислене за даною вибіркою, задовольняє нерівність $|t|<t_{двост,\alpha,k}$, то нульова гіпотеза H_0 приймається, вибірковий коефіцієнт кореляції r_B вважається статистично незначущим, величини X і Y некорельовані. У випадку $|t|>t_{двост,\alpha,k}$ ця гіпотеза відхиляється і приймається конкуруюча гіпотеза H_1, вибірковий коефіцієнт кореляції r_B вважається статистично значущим, величини X і Y корельовані.

Приклад 1. За вибіркою об'єму $n = 30$ з двовимірної генеральної сукупності (X, Y), що має нормальний розподіл, обчислено емпіричне значення $r_B = -0{,}68$ коефіцієнта кореляції r між X і Y. Необхідно при рівні значущості $\alpha = 0{,}02$ перевірити нульову гіпотезу $H_0: \ r = 0$ при альтернативній гіпотезі $H_1: \ r \neq 0$.

□ Обчислимо фактичне значення t статистики критерію T:

$$t = r_B \sqrt{(n-2)/(1-r_B^2)} = -0{,}68 \cdot \sqrt{(30-2)/(1-(-0{,}68)^2)} = -4{,}91.$$

За таблицею розподілу Стьюдента з $k = n - 2 = 30 - 2 = 28$ ступенями свободи при $\alpha = 0{,}02$ знайдемо двосторонню критичну точку: $t_{двост,\alpha,k} = t_{двост,0{,}02;28} = 2{,}47$. Оскільки

$$|t| = 4{,}91 > t_{двост,0{,}02;28} = 2{,}47,$$

то нульова гіпотеза $H_0: \ r = 0$ відхиляється. Тобто, при рівні значущості $\alpha = 0{,}02$ можна вважати, що емпіричний коефіцієнт кореляції r_B значущо відмінний від нуля, випадкові величини X і Y корельовані. ■

Зауваження. Якщо об'єм вибірки n малий, а емпіричне значення r_B коефіцієнта кореляції за модулем близьке до одиниці, то за статистику критерію необхідно прийняти випадкову величину $\boxed{T = \frac{\sqrt{n-3}}{2} \ln \frac{1+r_B}{1-r_B}}$, що при вірності нульової гіпотези H_0 також має розподіл Стьюдента з $k = n - 2$ ступенями свободи.

Приклад 2. За вибіркою об'єму $n = 7$ з двовимірної нормальної генеральної сукупності (X, Y) обчислено емпіричне значення $r_B = 0{,}96$ коефіцієнта кореляції r між X і Y. Необхідно при рівні значущості $\alpha = 0{,}01$ перевірити нульову гіпотезу $H_0: \ r = 0$ при альтернативній гіпотезі $H_1: \ r \neq 0$.

(Розв'язати самостійно. За статистику критерію прийняти $T = \left(\sqrt{n-3}/2\right) \ln\left((1+r_B)/(1-r_B)\right)$. Відповідь: гіпотеза H_0 приймається.)

2.7. Контрольні запитання до змістового модулю "Математична статистика"

1. Що таке математична статистика?
2. У чому полягає предмет математичної статистики?
3. Які основні задачі математичної статистики?
4. Що таке генеральна сукупність і вибіркова сукупність (вибірка)? У чому різниця між ними?
5. Яким умовам повинна задовольняти вибірка?
6. Опишіть способи відбору при формуванні вибірки.
7. Поясніть зміст вибіркового методу.
8. Що таке варіанта і варіаційний ряд?
9. Що таке частота і відносна частота варіанти? Емпірична щільність розподілу?
10. Що таке дискретний статистичний ряд? Інтервальний статистичний ряд?
11. Що таке полігон відносних частот і гістограма відносних частот?
12. Що таке емпірична функція розподілу і кумулятивна крива (кумулята)?
13. Що таке статистична оцінка параметра розподілу?
14. Чим відрізняються точкова й інтервальна оцінки параметра розподілу?
15. Що таке надійність (довірча ймовірність) і рівень значущості інтервальної оцінки?
16. Що служить незміщеною спроможною вибірковою оцінкою математичного сподівання?
17. Що служить спроможною вибірковою оцінкою дисперсії? У чому різниця між вибірковою дисперсією і виправленою вибірковою дисперсією?
18. Що таке вибіркове середнє квадратичне відхилення і виправлене вибіркове середнє квадратичне відхилення?
19. Дайте означення моди і медіани вибірки.
20. Як знаходять довірчі інтервали для математичного сподівання та дисперсії при різних припущеннях?
21. Що таке кореляційна таблиця?
22. Як для двовимірної випадкової величини (X,Y) визначають вибіркові оцінки: математичних сподівань і дисперсій, умовних математичних сподівань і дисперсій, коефіцієнта кореляції?

23. У чому полягає предмет і мета кореляційного аналізу?
24. В якому діапазоні лежать значення коефіцієнта кореляції?
25. З яких міркувань визначають тип кореляційної залежності $\bar{y}(x) = \varphi(x)$?
26. Що таке діаграма розсіювання (кореляційне поле) і емпірична (вибіркова) ламана регресії?
27. У чому суть методу найменших квадратів оцінки параметрів рівняння регресії?
28. Як одержують нормальну систему методу найменших квадратів?
29. Що таке статистична гіпотеза? Нульова й альтернативна (конкуруюча) гіпотези?
30. Що таке статистичний критерій перевірки гіпотез?
31. У чому полягає принцип практичної впевненості?
32. Поясніть значення термінів: статистика критерію, критична область, область прийняття гіпотези, критична точка.
33. В яких випадках вибирають лівосторонню, правосторонню чи двосторонню критичну область?
34. Які результати перевірки гіпотези відносять до помилок першого і другого роду?
35. Наведіть загальну схему критерію перевірки гіпотез.
36. Опишіть критерій перевірки гіпотези про математичне сподівання нормально розподіленої випадкової величини.
37. Коли на практиці виникає необхідність порівняння математичних сподівань? Як реалізується критерій перевірки гіпотези про рівність математичних сподівань двох нормально розподілених випадкових величин?
38. Коли на практиці виникає необхідність порівняння дисперсій? Опишіть критерій перевірки гіпотези про рівність дисперсій двох нормально розподілених випадкових величин
39. Що таке критерій згоди? Як здійснюється перевірка гіпотези про нормальний закон розподілу випадкової величини за критерієм Пірсона?
40. Як перевіряється гіпотеза про значущість коефіцієнта кореляції?

СПИСОК ЛІТЕРАТУРИ

1. Айвазян С.А., Енюков И.С., Мешалкин Л.Д. Прикладная статистика: Исследование зависимостей. – М.: Финансы и статистика, 1985. – 487 с.

2. Вентцель Е.С. Теория вероятностей и математическая статистика. – М.: Высш. школа, 1999. – 576 с.

3. Гмурман В.Е. Теория вероятностей и математическая статистика. – М.: Высш. шк., 2004. – 479 с.

4. Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике. – М.: Высш. шк., 2005. – 404 с.

5. Гусак А.А., Бричикова Е.А. Теория вероятностей: Справочное пособие к решению задач. – Мн.: НТООО “Тетра Системс”, 2000. – 288 с.

6. Жалдак М.І., Кузьміна Н.М., Берлінська С.Ю. Теорія ймовірностей і математична статистика з елементами інформаційної технології. – К.: Вища шк., 1995. – 351 с.

7. Жевняк Р.М., Карпук А.А., Унукович В.Т. Теория вероятностей и математическая статистика. – Мн.: Харвест, 2000. – 384 с.

8. Зубков А.М., Севастьянов Б.А., Чистяков В.П. Сборник задач по теории вероятностей. –М.: Наука, 1989. – 132 с.

9. Колде Я.К. Практикум по теории вероятностей и математической статистике. – М.: Высш. шк., 1991. – 158 с.

10. Копич І.М., Сороківський В.М. Елементи теорії ймовірностей і математичної статистики: теорія та практикум. – Львів: Вид-во ЛКА, 2001. – 336 с.

11. Кремер И.Ш. Теория вероятностей и математическая статистика. – М.: ЮНИТИ, 2006. – 573 с.

12. Методичні вказівки для практичних, самостійних та контрольних робіт з теорії ймовірностей та математичної статистики / Ю.Є. Печеніжський, С.О. Станішевський, В.С. Рухляда. – Х.: ХНАМГ, 2009. – 64 с.

13. Минько А.А. Статистический анализ в MS Excel. – М.: Изд. дом “Вильямс”, 2004. – 488 с.

14. Пак В.В., Носенко Ю.Л. Вища математика. – Донецьк: Сталкер, 2003. – 495 с.

15. Самойленко М.І., Кузнецов А.І., Костенко О.Б. Теорія ймовірностей. – Х.: ХНАМГ, 2008. – 194 с.

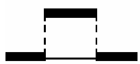
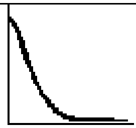
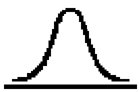
16. Справочник по теории вероятностей и математической статистике / Под ред. В.С. Королюка. – К.: Наук. думка, 1978. – 582 с.

17. Теорія імовірностей і математична статистика / А.Є. Ачкасов, В.Т. Плакіда та ін. – Х.: ХНАМГ, 2008. – 247 с.

ДОДАТКИ

Додаток 1

Найважливіші дискретні розподіли				
Назва розподілу	Ряд розподілу p_i	Параметри та їх можливі значення	Математичне сподівання M	Дисперсія D
Біноміальний	$C_n^i p^i q^{n-i}$, $i = 0,1,\dots,n$ $(q = 1 - p)$	n $(1,2,\dots)$; p $(0 \le p \le 1)$	np	npq
Пуассонівський	$a^i e^{-a} / i!$, $i = 0,1,\dots,n$	a $(a > 0)$	a	a

Найважливіші неперервні розподіли				
Назва розподілу	Щільність розподілу $f(x)$	Схема графіка $f(x)$	Математичне сподівання M	Дисперсія D
Рівномірний	$1/(\beta - \alpha)$, $\alpha \le x \le \beta$ і 0 в інших випадках		$\dfrac{\alpha + \beta}{2}$	$\dfrac{(\beta - \alpha)^2}{12}$
Показниковий	$\lambda e^{-\lambda x}$, $x > 0$ і 0 в інших випадках		$\dfrac{1}{\lambda}$	$\dfrac{1}{\lambda^2}$
Нормальний	$\dfrac{e^{-(x-a)^2/(2\sigma^2)}}{\sigma\sqrt{2\pi}}$, $-\infty < x < +\infty$		a	σ^2

Додаток 2

Значення функції Гаусса $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$

x	0	1	2	3	4	5	6	7	8	9
0,0	0,3989	3989	3989	3988	3986	3984	3982	3980	3977	3973
0,1	3970	3965	3961	3956	3951	3945	3939	3932	3925	3918
0,2	3910	3902	3894	3885	3876	3867	3857	3847	3836	3825
0,3	3814	3802	3790	3778	3765	3752	3739	3726	3712	3697
0,4	3683	3668	3653	3637	3621	3605	3589	3572	3555	3538
0,5	3521	3503	3485	3467	3448	3429	3410	3391	3372	3352
0,6	3332	3312	3292	3271	3251	3230	3209	3187	3166	3144
0,7	3123	3101	3079	3056	3034	3011	2989	2966	2943	2920
0,8	2897	2874	2850	2827	2803	2780	2756	2732	2709	2685
0,9	2661	2637	2613	2589	2565	2541	2516	2492	2468	2444
1,0	0,2420	2396	2371	2347	2323	2299	2275	2251	2227	2203
1,1	2179	2155	2131	2107	2083	2059	2036	2012	1989	1965
1,2	1942	1919	1895	1872	1849	1826	1804	1781	1758	1736
1,3	1714	1691	1669	1647	1626	1604	1582	1561	1539	1518
1,4	1497	1476	1456	1435	1415	1394	1374	1354	1334	1315
1,5	1295	1276	1257	1238	1219	1200	1182	1163	1145	1127
1,6	1109	1092	1074	1057	1040	1023	1006	0989	0973	0957
1,7	0940	0925	0909	0893	0878	0863	0848	0833	0818	0804
1,8	0790	0775	0761	0748	0734	0721	0707	0694	0681	0669
1,9	0656	0644	0632	0620	0608	0596	0584	0573	0562	0551
2,0	0,0540	0529	0519	0508	0498	0488	0478	0468	0459	0449
2,1	0440	0431	0422	0413	0404	0396	0387	0379	0371	0363
2,2	0355	0347	0339	0332	0325	0317	0310	0303	0297	0290
2,3	0283	0277	0270	0264	0258	0252	0246	0241	0235	0229
2,4	0224	0219	0213	0208	0203	0198	0194	0189	0184	0180
2,5	0175	0171	0167	0163	0158	0154	0151	0147	0143	0139

2,6	0136	0132	0129	0126	0122	0119	0116	0113	0110	0107
2,7	0104	0101	0099	0096	0093	0091	0088	0086	0084	0081
2,8	0079	0077	0075	0073	0071	0069	0067	0065	0063	0061
2,9	0060	0058	0056	0055	0053	0051	0050	0048	0047	0046
3,0	0,0044	0043	0042	0040	0039	0038	0037	0036	0035	0034
3,1	0033	0032	0031	0030	0029	0028	0027	0026	0025	0025
3,2	0024	0023	0022	0022	0021	0020	0020	0019	0018	0018
3,3	0017	0017	0016	0016	0015	0015	0014	0014	0013	0013
3,4	0012	0012	0012	0011	0011	0010	0010	0010	0009	0009
3,5	0009	0008	0008	0008	0008	0007	0007	0007	0007	0006
3,6	0006	0006	0006	0005	0005	0005	0005	0005	0005	0004
3,7	0004	0004	0004	0004	0004	0004	0003	0003	0003	0003
3,8	0003	0003	0003	0003	0003	0002	0002	0002	0002	0002
3,9	0002	0002	0002	0002	0002	0002	0002	0002	0001	0001

Додаток 3

Значення функції Лапласа $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-u^2/2} du$

x	0	1	2	3	4	5	6	7	8	9
0,0	0,00000	00399	00798	01197	01595	01994	02392	02790	03188	03586
0,1	03983	04380	04776	05172	05567	05962	06356	06749	07142	07535
0,2	07926	08317	08706	09095	09483	09871	10257	10642	11026	11409
0,3	11791	12172	12552	12930	13307	13683	14058	14431	14803	15173
0,4	15542	15910	16276	16640	17003	17364	17724	18082	18439	18793
0,5	19146	19497	19847	20194	20540	20884	21226	21566	21904	22240
0,6	22575	22907	23237	23565	23891	24215	24537	24857	25175	25490
0,7	25804	26115	26424	26730	27035	27337	27637	27935	28230	28524
0,8	28814	29103	29389	29673	29955	30234	30511	30785	31057	31327
0,9	31594	31859	32121	32381	32639	32894	33147	33398	33646	33891
1,0	0,34134	34375	34614	34850	35083	35314	35543	35769	35993	36214
1,1	36433	36650	36864	37076	37286	37493	37698	37900	38100	38298
1,2	38493	38686	38877	39065	39251	39435	39617	39796	39973	40147
1,3	40320	40490	40658	40824	40988	41149	41309	41466	41621	41774

1,4	41924	42073	42220	42364	42507	42647	42786	42922	43056	43189
1,5	43319	43448	43574	43699	43822	43943	44062	44179	44295	44408
1,6	44520	44630	44738	44845	44950	45053	45154	45254	45352	45449
1,7	45543	45637	45728	45818	45907	45994	46080	46164	46246	46327
1,8	46407	46485	46562	46638	46712	46784	46856	46926	46995	47062
1,9	47128	47193	47257	47320	47381	47441	47500	47558	47615	47670
2,0	0,47725	47778	47831	47882	47932	47982	48030	48077	48124	48169
2,1	48214	48257	48300	48341	48382	48422	48461	48500	48537	48574
2,2	48610	48645	48679	48713	48745	48778	48809	48840	48870	48899
2,3	48928	48956	48983	49010	49036	49061	49086	49111	49134	49158
2,4	49180	49202	49224	49245	49266	49286	49305	49324	49343	49361
2,5	49379	49396	49413	49430	49446	49461	49477	49492	49506	49520
2,6	49534	49547	49560	49573	49585	49598	49609	49621	49632	49643
2.7	49653	49664	49674	49683	49693	49702	49711	49720	49728	49736
2,8	49744	49752	49760	49767	49774	49781	49788	49795	49801	49807
2,9	49813	49819	49825	49831	49836	49841	49846	49851	49856	49861

x	3,0	3,1	3,2	3,3	3,4	3,5	3,6
$\Phi(x)$	0,49865	0,49903	0,49931	0,49952	0,49966	0,49977	0,49984
x	3,7	3,8	3,9	4,0	4,5	5,0	
$\Phi(x)$	0,49989	0,49993	0,49995	0,499968	0,499997	0,49999997	

Додаток 4

Критичні точки $t_{\alpha,k}$ t-розподілу (розподілу Стьюдента)

Число ступенів свободи k	Рівень значущості α (двостороння критична область $P(\|t\| > t_{двост,\alpha,k}) = \alpha$)					
	0,10	0,05	0,02	0,01	0,002	0,001
1	6,31	12,71	31,82	63,66	318,29	636,6
2	2,92	4,3	6,96	9,92	22,33	31,6
3	2,35	3,18	4,54	5,84	10,21	12,92
4	2,13	2,78	3,75	4,6	7,17	8,61
5	2,02	2,57	3,36	4,03	5,89	6,87

6	1,94	2,45	3,14	3,71	5,21	5,96
7	1,89	2,36	3	3,5	4,79	5,41
8	1,86	2,31	2,9	3,36	4,5	5,04
9	1,83	2,26	2,82	3,25	4,3	4,78
10	1,81	2,23	2,76	3,17	4,14	4,59
11	1,8	2,2	2,72	3,11	4,02	4,44
12	1,78	2,18	2,68	3,05	3,93	4,32
13	1,77	2,16	2,65	3,01	3,85	4,22
14	1,76	2,14	2,62	2,98	3,79	4,14
15	1,75	2,13	2,6	2,95	3,73	4,07
16	1,75	2,12	2,58	2,92	3,69	4,01
17	1,74	2,11	2,57	2,9	3,65	3,97
18	1,73	2,1	2,55	2,88	3,61	3,92
19	1,73	2,09	2,54	2,86	3,58	3,88
20	1,72	2,09	2,53	2,85	3,55	3,85
21	1,72	2,08	2,52	2,83	3,53	3,82
22	1,72	2,07	2,51	2,82	3,5	3,79
23	1,71	2,07	2,5	2,81	3,48	3,77
24	1,71	2,06	2,49	2,8	3,47	3,75
25	1,71	2,06	2,49	2,79	3,45	3,73
26	1,71	2,06	2,48	2,78	3,43	3,71
27	1,7	2,05	2,47	2,77	3,42	3,69
28	1,7	2,05	2,47	2,76	3,41	3,67
29	1,7	2,05	2,46	2,76	3,4	3,66
30	1,7	2,04	2,46	2,75	3,39	3,65
40	1,68	2,02	2,42	2,7	3,31	3,55
60	1,67	2	2,39	2,66	3,23	3,46
120	1,66	1,98	2,36	2,62	3,16	3,37
∞	1,64	1,96	2,33	2,58	3,09	3,29
	0,05	0,025	0,010	0,005	0,001	0,0005
	Рівень значущості α (одностороння критична область $P(\lvert t > t_{одност,\alpha,k}) = \alpha$)					

Додаток 5

Критичні точки $F_{кр,\alpha,k_1,k_2}$ F -розподілу

(розподілу Фішера – Снедекора) $P\left(F > F_{кр,\alpha,k_1,k_2}\right) = \alpha$

$\alpha = 0{,}05$								
k_2 \ k_1	2	3	4	5	6	8	12	24
2	19,00	19,16	19,25	19,30	19,33	19,37	19,41	19,45
3	9,55	9,28	9,12	9,01	8,94	8,84	8,74	8,64
5	5,79	5,41	5,19	5,05	4,95	4,82	4,68	4,53
6	5,14	4,76	4,53	4,39	4,28	4,15	4,00	3,84
8	4,46	4,07	3,84	3,69	3,58	3,44	3,28	3,12
10	4,10	3,71	3,48	3,33	3,22	3,07	2,91	2,74
12	3,88	3,49	3,26	3,11	3,00	2,85	2,69	2,50
15	3,68	3,29	3,06	2,90	2,79	2,64	2,48	2,29
17	3,59	3,20	2,96	2,81	2,70	2,55	2,38	2,19
20	3,49	3,10	2,87	2,71	2,60	2,45	2,28	2,08
25	3,38	2,99	2,76	2,60	2,49	2,34	2,16	1,96
30	3,32	2,92	2,69	2,53	2,42	2,27	2,09	1,89
40	3,23	2,84	2,62	2,45	2,34	2,18	2,00	1,79
60	3,15	2,76	2,52	2,37	2,25	2,10	1,92	1,70
120	3,07	2,68	2,45	2,29	2,17	2,02	1,83	1,61
$\alpha = 0{,}01$								
2	99,00	99,17	99,25	99,30	99,33	99,36	99,42	99,46
3	30,81	29,46	28,71	28,24	27,91	27,49	27,05	26,60
5	13,27	12,06	11,39	10,97	10,67	10,29	9,98	9,47

6	10,92	9,78	9,15	8,75	8,47	8,10	7,72	7,31
8	8,65	7,59	7,01	6,63	6,37	6,03	5,67	5,28
10	7,56	6,55	5,99	5,64	5,39	5,06	4,71	4,33
12	6,93	5,95	5,41	5,06	4,82	4,50	4,16	3,78
15	6,36	5,42	4,89	4,56	4,32	4,00	3,67	3,29
17	6,11	5,18	4,67	4,34	4,10	3,79	3,45	3,08
20	5,85	4,94	4,43	4,10	3,87	3,56	3,23	2,86
25	5,57	4,68	4,18	3,86	3,63	3,32	2,99	2,62
30	5,39	4,51	4,02	3,70	3,47	3,17	2,84	2,47
40	5,18	4,31	3,83	3,51	3,29	2,99	2,66	2,29
60	4,98	4,13	3,65	3,34	3,12	2,82	2,50	2,12
120	4,79	3,95	3,48	3,17	2,96	2,66	2,34	1,95

Додаток 6

Критичні точки $\chi^2_{кр,\alpha,k}$ χ^2-розподілу

(розподілу Пірсона) $P\left(\chi^2 > \chi^2_{кр,\alpha,k}\right) = \alpha$

Число ступенів свободи k	Рівень значущості α					
	0,01	0,025	0,05	0,95	0,975	0,99
1	6,6	5,0	3,8	0,0039	0,00098	0,00016
2	9,2	7,4	6,0	0,103	0,051	0,020
3	11,3	9,3	7,8	0,352	0,216	0,115
4	13,3	11,1	9,5	0,711	0,484	0,297
5	15,1	12,8	11,1	1,15	0,831	0,554

6	16,8	14,4	12,6	1,64	1,24	0,872
7	18,5	16,0	14,1	2,17	1,69	1,24
8	20,1	17,5	15,5	2,73	2,18	1,65
9	21,7	19,0	16,9	3,33	2,70	2,09
10	23,2	20,5	18,3	3,94	3,25	2,56
11	24,7	21,9	19,7	4,57	3,82	3,05
12	26,2	23,3	21,0	5,23	4,40	3,57
13	27,7	24,7	22,4	5,89	5,01	4,11
14	29,1	26,1	23,7	6,57	5,63	4,66
15	30,6	27,5	25,0	7,26	6,26	5,23
16	32,0	28,8	26,3	7,96	6,91	5,81
17	33,4	30,2	27,6	8,67	7,56	6,41
18	34,8	31,5	28,9	9,39	8,23	7,01
19	36,2	32,9	30,1	10,1	8,91	7,63
20	37,6	34,2	31,4	10,9	9,59	8,26
21	38,9	35,5	32,7	11,6	10,3	8,90
22	40,3	36,8	33,9	12,3	11,0	9,54
23	41,6	38,1	35,2	13,1	11,7	10,2
24	43,0	39,4	36,4	13,8	12,4	10,9
25	44,3	40,6	37,7	14,6	13,1	11,5
26	45,6	41,9	38,9	15,4	13,8	12,2
27	47,0	43,2	40,1	16,2	14,6	12,9
28	48,3	44,5	41,3	16,9	15,3	13,6
29	49,6	45,7	42,6	17,7	16,0	14,3
30	50,9	47,0	43,8	18,5	16,8	15,0

З М І С Т

НАВЧАЛЬНЕ ВИДАННЯ

Колосов Анатолій Іванович,
Печеніжський Юрій Євгенович,
Станішевський Степан Олександрович,
Якунін Анатолій Вікторович

ТЕОРІЯ ЙМОВІРНОСТЕЙ і МАТЕМАТИЧНА СТАТИСТИКА

КОНСПЕКТ ЛЕКЦІЙ
(для студентів 2 курсу заочної форми навчання
за напрямами підготовки 6.030504 „Економіка підприємства”
і 6.030509 “Облік і аудит”)

Відповідальний за випуск *Л. Б. Коваленко*
Редактор *З. І. Зайцева*
Комп’ютерне верстання *А. В. Якунін*

План 2011, поз. 97Л

Підп. до друку 28.11.2011 Формат 60х84 1/16
Друк на ризографі Ум. друк. арк. 7,0
Тираж 100 пр. Зам. №

Видавець і виготовлювач:
Харківська національна академія міського господарства,
вул. Революції, 12, Харків, 61002
Електронна адреса: rectorat@ksame.kharkov.ua
Свідоцтво суб’єкта видавничої справи:
ДК № 4064 від 12.05.2011